Dispersion analysis of compatible Galerkin schemes for the 1D shallow water model

Christopher Eldred, Daniel Le Roux

Dispersion analysis of compatible Galerkin schemes for the 1D shallow water model

Christopher Eldred & Daniel Y. Le Roux*

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France*

Abstract

In this work, we study the dispersion properties of two compatible Galerkin schemes for the 1D linearized shallow water equations: the $P_n^C - P_{n-1}^{DG}$ and the $GD_n - DGD_{n-1}$ element pairs. P_n is the order n Lagrange space, P_{n-1}^{DG} is the order $n-1$ discontinuous Lagrange space, GD_n is the order n Galerkin difference space, and DGD_{n-1} is the order $n-1$ discontinuous Galerkin difference space. Compatible Galerkin methods have many desirable properties, including energy conservation, steady geostrophic modes and the absence of spurious stationary modes, such as pressure modes. However, this does not guarantee good wave dispersion properties. Previous work on the $P_2^C - P_1^{DG}$ pair has indeed indicated the presence of spectral gaps, and it is extended in this paper to the study of the $P_n^C - P_{n-1}^{DG}$ pair for arbitrary n. Additionally, an alternative element pair is introduced, the $GD_n - DGD_{n-1}$ pair, that is free of spectral gaps while benefiting from the desirable properties of compatible elements. Asymptotic convergence rates are established for both element pairs, including the use of inexact quadrature (which diagonalizes the velocity mass matrix) for the $P_n^C - P_{n-1}^{DG}$ pair and reduced quadrature for the $GD_n - DGD_{n-1}$ pair. Plots of the dispersion relationship and group velocities for a wide range of n and Rossby radii are shown. A brief investigation into the utility of mass lumping to remove the spectral gaps for the $P_3^C - P_2^{DG}$ pair is performed. Finally, a pair of numerical simulations are run to investigate the consequences of the spectral gaps and highlight the main differences between the two elements.

Keywords: dispersion relationship, shallow water equations, geophysical fluid dynamics, dynamical core, mixed finite elements, finite element

*Corresponding Author: Christopher Eldred (chris.eldred@gmail.com)

1. Introduction

The application of compact Galerkin methods for numerical models of geophysical fluid flows has become increasingly common over the last 15 years. Of particular interest are compatible finite element methods [1, 2] or their closely related generalization compatible Galerkin methods [3]. Compatible Galerkin methods are the generalization of discrete deRham complexes from finite element exterior calculus [4] to more general compact Galerkin methods that are not finite elements. Some compatible finite element pairs were first investigated in [5, 6]. Such pairs represent the extension of Arakawa C-grid finite difference schemes (also known as staggered grid method or Marker and Cell methods) to a Galerkin approach. Due to this, they have many desirable properties, such as energy conservation, steady geostrophic modes and various mimetic properties. Additionally, compatible Galerkin methods do not suffer from spurious pressure modes or inertial modes, although all known examples do have a CD/Coriolis mode [1, 7] due to the discrete Coriolis matrix being rank-deficient. However, even then compatible Galerkin schemes are not guaranteed to have good wave dispersion properties. Further care in the choice of spaces is required to ensure a lack of spurious branches and the appearance of spectral gaps in the dispersion relationship. A precise definition for branches, spurious branches and spectral gaps will be given in Section 3. The dispersion properties of compatible Galerkin schemes for the 1D linear shallow water equations is the focus of this paper.

The history of dispersion analysis goes back to the dawn of geophysical fluid modeling [8]. However, only recently have the dispersion properties of compatible finite element methods for the shallow water model been investigated. The $RT_0 - P_0^{DG}$ and/or $BDM_1 - P_0^{DG}$ pairs on triangles were studied in [5, 6, 7, 9, 10]. These are the lowest-order members of the corresponding $P_r^- \Lambda^k$ (RT_0) and $P_r \Lambda^k$ (BDM_1) families from finite element exterior calculus [4]. Unfortunately both elements have spurious branches of the dispersion relationship, and the presence or absence of spectral gaps is still unclear. Such gaps are unphysical numerical artifacts, and are a general feature of high order finite element discretizations [11, 12, 13]. The compound finite elements (which are a compatible finite element method on arbitrary polygons) introduced in [14] were studied in [15] for quadrilaterals and hexagons,

and found to be quite similar to the corresponding C-grid finite difference schemes. The $P_2^C - P_1^{DG}$ pair in 1D was analyzed in [16], and a spectral gap was found. A solution to this gap for the nonrotating linear shallow water equations, obtained through partial lumping of the velocity mass matrix, is given in [17]. This approach was extended to the $RT_1 - P_1^{DG}$ pair on quadrilaterals for the 2D rotating linear shallow water equations in [18]. When this work was started, detailed study of the dispersion properties of the $P_n^C - P_{n-1}^{DG}$ pair for arbitrary n was still lacking, and was one of the focuses of this work. However, late in the process of preparing this manuscript, we became aware of a related, independent study for the $P_n^C - P_{n-1}^{DG}$ element [19], which also studied the 2D dispersion relationship for pure gravity waves. This work extends [19] in several keys ways: a focus on inertia-gravity waves rather than the limiting cases of pure gravity ($f = 0$) or pure inertia ($g = H = 0$) waves; determination of the asymptotics for the $P_n^C - P_{n-1}^{DG}$ pair with both exact and inexact quadrature, and also for the partially mass lumped version from [17, 18]; and an alternative branch selection procedure (see Section 3.5).

In this paper, the dispersion properties of two compatible Galerkin schemes are studied for the 1D linearized shallow water equations: the $P_n^C - P_{n-1}^{DG}$ and the $GD_n - DGD_{n-1}$ finite element pairs. P_n is the order n Lagrange space, P_{n-1}^{DG} is the order $n-1$ discontinuous Lagrange space, GD_n is the order n Galerkin difference space, and DGD_{n-1} is the order $n-1$ discontinuous Galerkin difference space. More details on these two pairs are provided in Section 4 and Section 5. We show that the number and width of the spectral gaps for the $P_n^C - P_{n-1}^{DG}$ pair increases as n increases. The $P_n^C - P_{n-1}^{DG}$ pair is investigated for both exact and inexact quadrature, with inexact leading to a diagonal velocity mass matrix. In addition, partial lumping of the velocity mass matrix is explored as a means to eliminate the spectral gaps that occur for $n \geq 2$. The presence of such gaps motivates the introduction of the $GD_n - DGD_{n-1}$ pair that does not suffer from spectral gaps. For the $GD_n - DGD_{n-1}$ pair, both exact and reduced quadrature are investigated, with reduced quadrature ameliorating some of the computational cost, although the mass matrix remains non-diagonal. Finally, both schemes are compared using a range of n for two test cases, and some conclusions about their applicability to the development of a full geophysical fluid model are drawn. This work represents a starting point for the analysis of the dispersion properties of compatible Galerkin schemes applied to the shallow water equations: in particular, an extension to 2D inertia-gravity waves and the incorporation of time discretization remains to be done.

2. Model Problem

Consider the 1D, inviscid shallow water (SW) equations with constant Coriolis parameter f and a flat bottom, linearized about a state of rest with constant fluid depth H. Such a formulation is satisfactory for our purpose, which in Cartesian coordinates is expressed [20] as

$$\frac{\partial u}{\partial t} - f\,v + g\,\frac{\partial \eta}{\partial x} = 0\,, \tag{1}$$

$$\frac{\partial v}{\partial t} + f\,u = 0\,, \tag{2}$$

$$\frac{\partial \eta}{\partial t} + H\,\frac{\partial u}{\partial x} = 0\,, \tag{3}$$

where $\mathbf{u}(x,t) = (u,v)$ is the velocity component, $\eta(x,t)$ is the surface elevation with respect to the reference level $z = 0$ and g is the gravitational acceleration. Note that η would be the pressure in the Navier–Stokes equations. Equations (1) - (3) describe a first order hyperbolic system, and initial conditions and periodic boundary conditions (which are employed for the subsequent Fourier analyses) complete the mathematical statement of the problem.

The linear stability of (1) - (3) is first examined. Because (1) - (3) form a linear system with constant coefficients in a periodic domain, it is sufficient to consider a single Fourier mode, and the variables can be written as $\eta = \hat{\eta}e^{-i\omega t}e^{ikx}$, $u = \hat{u}e^{-i\omega t}e^{ikx}$ and $v = \hat{v}e^{-i\omega t}e^{ikx}$, where $\hat{\eta}$, $\hat{u}$ and $\hat{v}$ are the Fourier amplitudes, ω is the temporal frequency and k is the wavenumber. Substitution into (1) - (3) leads to

$$-i\omega\hat{u} - f\hat{v} + ikg\hat{\eta} = 0, \tag{4}$$

$$-i\omega\hat{v} + f\hat{u} = 0, \tag{5}$$

$$-i\omega\hat{\eta} + ikH\hat{u} = 0. \tag{6}$$

The system (4) - (6) can be solved for non-trivial ($\hat{\eta}$, $\hat{u}$, $\hat{v}$) and gives a dispersion polynomial of degree 3 in ω. The solutions are the geostrophic mode $\omega = 0$ and two analytical (AN) inertia-gravity modes

$$\omega_{AN} = \pm\sqrt{f^2 + gHk^2}. \tag{7}$$

Equation (7) is rewritten in the form

$$\sigma^2 = 1 + \lambda^2 k^2, \tag{8}$$

where $\sigma = \frac{\omega_{AN}}{f}$ is the non-dimensional temporal frequency and $\lambda = \frac{\sqrt{gH}}{f}$ is the Rossby radius of deformation. Note that ω_{AN} and σ are both real. To facilitate comparison with numerical results, it useful to rewrite (8) as

$$\sigma^2 = 1 + \mu(kh)^2, \tag{9}$$

where h is some measure of length (taken later on to equal to the element width), $kh \in [0, \pi]$ is the non-dimensional wavenumber and $\mu = \frac{\lambda^2}{h^2}$ is the (squared) non-dimensional ratio of Rossby radius of deformation to element width. The group velocity can be obtained from (9)

$$\frac{\partial \sigma}{\partial (kh)} = \pm \frac{\mu kh}{(1 + \mu(kh)^2)^{\frac{1}{2}}}. \tag{10}$$

Both the group velocity and the dispersion relation are monotonic and non-zero for all $kh > 0$, provided g, f and H are all non-zero. We do not consider the limiting cases of $f = 0$ (pure gravity waves) or $g = H = 0$ (pure inertia waves) in this paper. They are considered for the $P_n^C - P_{n-1}^{DG}$ pair in [19].

3. General Compatible Galerkin Discretization in 1D

In what follows, we assume continuity in time. Start by spatially discretizing (1) - (3) using a Galerkin approach: assume that the prognostic variables η, u and v belong to some function space, multiply each equation by a test function from that function space, integrate over the domain, and require that these equations hold for every possible choice of test function. Following the approach of finite element exterior calculus [4], we let $\eta, v \in \mathbb{B} \subset L_2$ and $u \in \mathbb{A} \subset H^1$, such that $\mathbb{A}$ and $\mathbb{B}$ form an exact sequence under the $\frac{d}{dx}$ operator, which is a discrete version of the 1D deRham complex. In the finite element literature, this approach is also known as mixed finite elements, since different function spaces are used for the various fields. From these 1D spaces, a family of spaces that build a discrete deRham complex in any number of dimensions can be generated by taking appropriate tensor products of the 1D elements. Denoting the test functions by $\eta^*, v^* \in \mathbb{B}$, $u^* \in \mathbb{A}$ this

gives, after appropriate integration by parts,

$$\left\langle u^*, \frac{\partial u}{\partial t} \right\rangle - f \left\langle u^*, v \right\rangle - g \left\langle \frac{\partial}{\partial x} u^*, \eta \right\rangle = 0, \tag{11}$$

$$\left\langle v^*, \frac{\partial v}{\partial t} \right\rangle + f \left\langle v^*, u \right\rangle = 0, \tag{12}$$

$$\left\langle \eta^*, \frac{\partial \eta}{\partial t} \right\rangle + H \left\langle \eta^*, \frac{\partial}{\partial x} u \right\rangle = 0, \tag{13}$$

where $\langle , \rangle$ is the L_2 inner product. Writing (11) - (13) in matrix form yields

$$\mathbf{M_u} \frac{\partial \mathbf{u}}{\partial t} - f \mathbf{C_u v} + g \mathbf{G} \boldsymbol{\eta} = 0, \tag{14}$$

$$\mathbf{M_v} \frac{\partial \mathbf{v}}{\partial t} + f \mathbf{C_v u} = 0, \tag{15}$$

$$\mathbf{M}_{\boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}}{\partial t} + H \mathbf{D u} = 0, \tag{16}$$

where $\mathbf{u}$, $\mathbf{v}$ and $\boldsymbol{\eta}$ denote the basis coefficients in the Galerkin expansion and the matrices are given by

$$\mathbf{M_u} \frac{\partial \mathbf{u}}{\partial t} = \left\langle u^*, \frac{\partial u}{\partial t} \right\rangle, \quad \mathbf{M_v} \frac{\partial \mathbf{v}}{\partial t} = \left\langle v^*, \frac{\partial v}{\partial t} \right\rangle, \quad \mathbf{M}_{\boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}}{\partial t} = \left\langle \eta^*, \frac{\partial \eta}{\partial t} \right\rangle$$

$$\mathbf{G}\boldsymbol{\eta} = - \left\langle \frac{\partial}{\partial x} u^*, \eta \right\rangle, \quad \mathbf{Du} = \left\langle \eta^*, \frac{\partial}{\partial x} u \right\rangle, \quad \mathbf{C_u v} = \left\langle u^*, v \right\rangle, \quad \mathbf{C_v u} = \left\langle v^*, u \right\rangle.$$

The choice of function spaces implies

$$\mathbf{M_u} = \mathbf{M_u}^*, \tag{17}$$

$$\mathbf{M_v} = \mathbf{M}_{\boldsymbol{\eta}} = \mathbf{M_v}^* = \mathbf{M}_{\boldsymbol{\eta}}^*, \tag{18}$$

$$\mathbf{G}^* = -\mathbf{D}, \tag{19}$$

$$\mathbf{C_u}^* = \mathbf{C_v}. \tag{20}$$

where $\mathbf{A}^*$ denotes the Hermitian transpose. Note that (19) and (20) ensure that the pressure-gradient and Coriolis terms are energetically neutral, respectively. These properties will be useful later on, in proving some general features of the discrete dispersion relationship.

3.1. Fourier Analysis

From now on, consider a periodic computational grid made of N uniform elements of width h. For what follows, l is an integer that depends on the choice of spaces, and is discussed further in Section 4 and Section 5. The matrices in (14) - (16) are of size $Nl \times Nl$ with $Nl = N \times l$. Now assume that the basis coefficients have a Fourier dependence in space and time, as done in the continuous case:

$$u_j(t) = \tilde{u}_j e^{-i\omega t} e^{ikx_j}, \quad v_j(t) = \tilde{v}_j e^{-i\omega t} e^{ikx_j}, \quad \eta_j(t) = \tilde{\eta}_j e^{-i\omega t} e^{ikx_j}, \tag{21}$$

where $\tilde{u}_j$, $\tilde{v}_j$, $\tilde{\eta}_j$, $j = 1, \ldots, Nl$, are the Fourier amplitudes for the jth basis coefficient, and x_j is the nodal location of the jth basis function for the relevant variable. Inserting (21) into (14) - (16) yields

$$-i\omega\widetilde{\mathbf{M}}_{\mathbf{u}}\tilde{\mathbf{u}} - f\widetilde{\mathbf{C}}_{\mathbf{u}}\tilde{\mathbf{v}} + g\widetilde{\mathbf{G}}\tilde{\boldsymbol{\eta}} = 0, \tag{22}$$

$$-i\omega\widetilde{\mathbf{M}}_{\mathbf{v}}\tilde{\mathbf{v}} + f\widetilde{\mathbf{C}}_{\mathbf{v}}\tilde{\mathbf{u}} = 0, \tag{23}$$

$$-i\omega\widetilde{\mathbf{M}}_{\boldsymbol{\eta}}\tilde{\boldsymbol{\eta}} + H\widetilde{\mathbf{D}}\tilde{\mathbf{u}} = 0, \tag{24}$$

where

$$\tilde{\mathbf{u}} = [\tilde{u}_1, \ldots, \tilde{u}_{Nl}], \quad \tilde{\mathbf{v}} = [\tilde{v}_1, \ldots, \tilde{v}_{Nl}], \quad \tilde{\boldsymbol{\eta}} = [\tilde{\eta}_1, \ldots, \tilde{\eta}_{Nl}]. \tag{25}$$

The relationship between any matrix $\mathbf{A}$ and $\widetilde{\mathbf{A}}$ is given by

$$\widetilde{\mathbf{A}}_{mn} = \mathbf{A}_{mn} e^{ik(x_m - x_n)}, \tag{26}$$

where we have normalized each row by e^{ikx_m} (with x_m the location of the degree of freedom for row m). Finally, utilizing translational invariance, there are actually only l unique degrees of freedom for $\tilde{\mathbf{u}}$, $\tilde{\mathbf{v}}$ and $\tilde{\boldsymbol{\eta}}$:

$$\hat{\mathbf{u}} = [\hat{u}_1, \ldots, \hat{u}_l]^T, \quad \hat{\mathbf{v}} = [\hat{v}_1, \ldots, \hat{v}_l]^T, \quad \hat{\boldsymbol{\eta}} = [\hat{\eta}_1, \ldots, \hat{\eta}_l]^T. \tag{27}$$

Therefore,

$$\tilde{\mathbf{u}} = [\hat{u}_1, \ldots, \hat{u}_l, \hat{u}_1, \ldots, \hat{u}_l, \ldots], \tag{28}$$

and similarly for $\tilde{\mathbf{v}}$ and $\tilde{\boldsymbol{\eta}}$. Considering now a single element (say element $\frac{N}{2}$) and only the l rows for the degrees of freedom associated with this element, we can write the final discrete system as

$$-i\omega\widehat{\mathbf{M}}_{\mathbf{u}}\hat{\mathbf{u}} - f\widehat{\mathbf{C}}_{\mathbf{u}}\hat{\mathbf{v}} + g\widehat{\mathbf{G}}\hat{\boldsymbol{\eta}} = 0, \tag{29}$$

$$-i\omega\widehat{\mathbf{M}}_{\mathbf{v}}\hat{\mathbf{v}} + f\widehat{\mathbf{C}}_{\mathbf{v}}\hat{\mathbf{u}} = 0, \tag{30}$$

$$-i\omega\widehat{\mathbf{M}}_{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}} + H\widehat{\mathbf{D}}\hat{\mathbf{u}} = 0. \tag{31}$$

The matrices $\widehat{\mathbf{A}}$ and $\widetilde{\mathbf{A}}$ are related by

$$\hat{\mathbf{A}}_{mn} = \sum_{p} \tilde{\mathbf{A}}_{mp}, \tag{32}$$

with $m = 1, \ldots, l$, $n = 1, \ldots, l$, and p is the set of integers belonging to the interval $[1, \ldots, Nl]$ such that $\tilde{x}_p = \hat{x}_n$. Note that the matrices keep their properties through these transformations, which is essential to proving various properties of the dispersion relationship. System (29) - (31) is rewritten in the form of a generalized eigenvalue problem (GEP)

$$-i\omega\widehat{\mathbf{M}}\hat{\mathbf{x}} = \widehat{\mathbf{S}}\hat{\mathbf{x}}, \tag{33}$$

where $\hat{\mathbf{x}} = [\hat{u}_1, \ldots, \hat{u}_l, \hat{v}_1, \ldots, \hat{v}_l, \hat{\eta}_1, \ldots, \hat{\eta}_l]^T$ and

$$\widehat{\mathbf{M}} = \begin{pmatrix} \widehat{\mathbf{M}}_{\mathbf{u}} & 0 & 0 \\ 0 & \widehat{\mathbf{M}}_{\mathbf{v}} & 0 \\ 0 & 0 & \widehat{\mathbf{M}}_{\boldsymbol{\eta}} \end{pmatrix}, \quad \widehat{\mathbf{S}} = \begin{pmatrix} 0 & f\widehat{\mathbf{C}}_{\mathbf{u}} & -g\widehat{\mathbf{G}} \\ -f\widehat{\mathbf{C}}_{\mathbf{v}} & 0 & 0 \\ -H\widehat{\mathbf{D}} & 0 & 0 \end{pmatrix}. \tag{34}$$

Both $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{S}}$ are 3×3 block matrices of total size $3l \times 3l$, with individual blocks of size $l \times l$. Specific expressions for the individual matrix blocks and the discrete degrees of freedom $\hat{\mathbf{x}}$ will depend on the choice of spaces $\mathbb{A}$ and $\mathbb{B}$, as detailed in Section 4 and Section 5. The solutions to (33) are the linear modes of the discretized system, which form the dispersion relationship.

Proposition 1. *The generalized eigenvalue problem (33) has $3l$ purely imaginary eigenvalues (and thus ω is real): l geostrophic modes with $\omega = 0$ and $2l$ inertia-gravity waves. Additionally, the solution of (33) can be put into the following reduced form using block matrix determinants*

$$det(-\omega^2\widehat{\mathbf{M}}_{\mathbf{u}} - gH\widehat{\mathbf{G}}(\widehat{\mathbf{M}}_{\boldsymbol{\eta}})^{-1}\widehat{\mathbf{D}} + f^2\widehat{\mathbf{C}}_{\mathbf{u}}(\widehat{\mathbf{M}}_{\mathbf{v}})^{-1}\widehat{\mathbf{C}}_{\mathbf{v}}) := P_n(\sigma), \tag{35}$$

where $P_n(\sigma)$ is the characteristic polynomial of degree l in ω^2, which is called the dispersion polynomial in the following.

Proof. Starting with (33), and using (17) - (20), it is clear that

$$\widehat{\mathbf{M}} = \widehat{\mathbf{M}}^*, \tag{36}$$

and

$$\widehat{\mathbf{S}} = -\widehat{\mathbf{S}}^*. \tag{37}$$

Therefore, the generalized eigenvalue problem (33) will have purely imaginary eigenvalues, and thus ω is real. This is another way of stating linear energy conservation.

To show (35), start by writing (33) in the form

$$(\widehat{\mathbf{S}} + i\omega\widehat{\mathbf{M}})\hat{\mathbf{x}} = \widehat{\mathbf{F}}\hat{\mathbf{x}} = 0 \tag{38}$$

where $\widehat{\mathbf{F}} = \widehat{\mathbf{S}} + i\omega\widehat{\mathbf{M}}$ and the solutions therefore satisfy

$$\det(\widehat{\mathbf{F}}) = 0 \tag{39}$$

Now we partition $\widehat{\mathbf{F}}$ as a block matrix

$$\widehat{\mathbf{F}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \tag{40}$$

with

$$\mathbf{A}_{11} = i\omega\widehat{\mathbf{M}}_{\mathbf{u}} \quad \mathbf{A}_{12} = \begin{pmatrix} f\widehat{\mathbf{C}}_{\mathbf{u}} & -g\widehat{\mathbf{G}} \end{pmatrix} \quad \mathbf{A}_{21} = \begin{pmatrix} -f\widehat{\mathbf{C}}_{\mathbf{v}} \\ -H\widehat{\mathbf{D}} \end{pmatrix} \quad \mathbf{A}_{22} = \begin{pmatrix} i\omega\widehat{\mathbf{M}}_{\mathbf{v}} & 0 \\ 0 & i\omega\widehat{\mathbf{M}}_{\boldsymbol{\eta}} \end{pmatrix} \tag{41}$$

Using the Schur complement, the determinant of $\widehat{\mathbf{F}}$ can be written as

$$\det(\widehat{\mathbf{F}}) = \det(\mathbf{A}_{22})\det(\mathbf{A}_{11} - \mathbf{A}_{12}(\mathbf{A}_{22})^{-1}\mathbf{A}_{21}) \tag{42}$$

since $\mathbf{A}_{22}$ is invertible. Plugging in (41) yields

$$\begin{aligned} \det(\widehat{\mathbf{F}}) = \det(i\omega\widehat{\mathbf{M}}_{\mathbf{v}})\det(i\omega\widehat{\mathbf{M}}_{\boldsymbol{\eta}}) \\ \det(i\omega\widehat{\mathbf{M}}_{\mathbf{u}} + \frac{gHi}{\omega}\widehat{\mathbf{G}}(\widehat{\mathbf{M}}_{\boldsymbol{\eta}})^{-1}\widehat{\mathbf{D}} - \frac{f^2 i}{\omega}\widehat{\mathbf{C}}_{\mathbf{u}}(\widehat{\mathbf{M}}_{\mathbf{v}})^{-1}\widehat{\mathbf{C}}_{\mathbf{v}}) = 0, \end{aligned} \tag{43}$$

which can be simplified to give

$$\omega^l \det(-\omega^2\widehat{\mathbf{M}}_{\mathbf{u}} - gH\widehat{\mathbf{G}}(\widehat{\mathbf{M}}_{\boldsymbol{\eta}})^{-1}\widehat{\mathbf{D}} + f^2\widehat{\mathbf{C}}_{\mathbf{u}}(\widehat{\mathbf{M}}_{\mathbf{v}})^{-1}\widehat{\mathbf{C}}_{\mathbf{v}}) = 0. \tag{44}$$

The first solution is the geostrophic mode, corresponding to $\omega = 0$, which has multiplicity l. The remaining $2l$ eigenvalues, which are the inertia-gravity modes, are determined by solving the dispersion polynomial $P_n(\omega)$ associated with the determinant of $\hat{\mathbf{L}}$:

$$\hat{\mathbf{L}} = -\omega^2\widehat{\mathbf{M}}_{\mathbf{u}} - gH\widehat{\mathbf{G}}(\widehat{\mathbf{M}}_{\boldsymbol{\eta}})^{-1}\widehat{\mathbf{D}} + f^2\widehat{\mathbf{C}}_{\mathbf{u}}(\widehat{\mathbf{M}}_{\mathbf{v}})^{-1}\widehat{\mathbf{C}}_{\mathbf{v}}. \tag{45}$$

This is just another way of writing (35). These solutions come in pairs due to the presence of ω^2. Using the reduced form (35) instead of (33) leads to a dispersion polynomial of degree l instead of $3l$, which is much easier to solve when l is large. □

3.2. Non-Dimensional Wavenumbers

It is useful to define an adjusted grid spacing

$$\tilde{h} = \frac{h}{l}, \tag{46}$$

that is the distance between finite element degrees of freedom (nodes) for equally spaced nodes. However, in practice non-uniformly spaced nodes are often used (to improve the conditioning of the resulting matrices), and so $\tilde{h}$ can be though of as an average distance. Using $\tilde{h}$, a non-dimensional wavenumber is naturally defined as

$$\widetilde{kh} = k\tilde{h}, \tag{47}$$

with the associated wavelength

$$\zeta = \frac{2\pi}{k} = \frac{2\pi\tilde{h}}{\widetilde{kh}}, \tag{48}$$

and non-dimensional wavelength

$$\tilde{\zeta} = \frac{\zeta}{\tilde{h}} = \frac{2\pi}{\widetilde{kh}}. \tag{49}$$

For a grid with N elements, the maximal wavelength is $\zeta = Nl\tilde{h} = Nh$, with $k = \frac{2\pi}{Nh}$, $\widetilde{kh} = \frac{2\pi}{Nl}$ and $\tilde{\zeta} = Nl$; and the minimum wavelength is $\zeta = 2\tilde{h}$, with $k = \frac{\pi}{\tilde{h}}$, $\widetilde{kh} = \pi$ and $\tilde{\zeta} = 2$.

3.3. Branches, Spurious Branches and Spectral Gaps

In general, the dispersion polynomial $P_n(\sigma)$ in (35) will have l pairs of solutions corresponding to inertia-gravity modes. When $l \geq 2$, for a given non-dimensional wavenumber $\widetilde{kh}$, only one of these solutions should be retained in order to correspond to the analytic dispersion relationship. Therefore, for each solution of $P_n(\sigma)$ there exists a non-dimensional wavenumber range over which it is retained, which is termed a branch in this paper. The union of all branches constitutes the discrete dispersion relationship. The remaining parts of each solution are mathematical artifacts, and should be disregarded. It is key to note here that for a given non-dimensional wavenumber, there is only a single propagating wave, not a collection of waves. A graphical

demonstration is obtained by looking at any of the figures plotting dispersion relationships and group velocities for the $P_n^C - P_{n-1}^{DG}$ pair. For example, consider Figure 4 and look at the $n = 2$ row. In the left column are the solutions of the dispersion polynomial $P_2(\sigma)$, with the two solutions colored blue and red. In the middle column is the dispersion relationship, which consists of two branches, each of which is a portion of a solution from the dispersion polynomial. Each branch is colored according to the solution from which it came. The right column has the group velocity, using the same coloring scheme. When $l = 1$, there is only one solution and it is retained over the whole wavenumber range. At the boundaries between two branches, the dispersion relationship is usually discontinuous. This defines a spectral gap as a spatial wavenumber for which the dispersion relationship becomes double valued and the group velocity goes to zero. These gaps result in a piecewise continuous dispersion relationship. In the continuous case, for every geostrophic mode there are 2 inertia-gravity modes. This is also the case for compatible Galerkin methods in 1D, but it is not always the case in 2D. For example, the $RT_0 - P_0^{DG}$ and $BDM_1 - P_0^{DG}$ pairs on triangles do not satisfy this 2:1 ratio. We define an element pair as possessing spurious branches of the dispersion relationship if there are not two inertia-gravity modes for every geostrophic mode (or Rossby mode in the case of variable f). The exact nature of the spurious modes, when they exist, their relationship to spectral gaps and the way of assembling the different branches over the whole spectrum to produce the discrete dispersion relationship; especially on triangular elements; is still not clear.

3.4. Solution Procedure

To solve (35), two complementary approaches are used. The first approach uses a computer algebra system (Maple) to analytically construct the matrices and solve the determinant to obtain the dispersion relationship $\sigma(kh, \mu)$. A series expansion around $h = 0$ is then performed in order to obtain the asymptotics of the dispersion relationship. At higher l this procedure is found to be quite expensive, due to the complicated nature of the analytic expressions and the need to compute the determinant of an $l \times l$ matrix and solve the resulting polynomial of degree l.

For the second approach, a computational grid of size $N = \frac{6000}{l}$ is used. The eigenvalues and eigenvectors of (35) are determined numerically for each kh and μ of interest. The complete set of results obtained constitutes the dispersion polynomial $P_n(\sigma)$, which is used to plot the dispersion relationship,

group velocity and effective resolution. The set of kh sampled is a uniformly spaced set belonging to $[\frac{2\pi}{Nl}, \pi]$. When $l \geq 2$ (which occurs for the $P_n^C - P_{n-1}^{DG}$ pair with $n \geq 2$), for each kh there will be l eigenvalues, which are the l solutions of the dispersion polynomial. However, it turns out that only one solution is physical, and the others are mathematical artifacts. The following section outlines the method used to select the physical solution.

3.5. Branch determination for $l \geq 2$

The following procedure is used to select the appropriate branch:

1. Let $kh \in (0, \pi]$ and numerically compute the l eigenvalues ω_m ($m = 1, \ldots, l$) and eigenvectors $\hat{x}_m$.
2. For each eigenvalue and eigenvector pair, determine the branch p that it belongs to. This is done by constructing the discrete solution corresponding to $\hat{x}_m$ and comparing it to the continuous solution e^{ikx}.
3. Determine the effective wavenumber $\widetilde{kh}$ associated with ω_p using the explicit formula
$$\widetilde{kh} = \frac{(-1)^\alpha (kh) + 2\pi\beta}{l}, \tag{50}$$
where $\alpha = 1 + p$ and $\beta = \lfloor \frac{p}{2} \rfloor$. This procedure is equivalent to the mode association done in [11], and it gives an effective wavenumber range of $(0, \pi]$.

An example for $l = 2$ is provided in Section 4.1.

4. Finite Element Exterior Calculus: the $P_n^C - P_{n-1}^{DG}$ pair

One possible choice for the $\mathbb{A}$ and $\mathbb{B}$ spaces is $\mathbb{A} = P_n^C$ and $\mathbb{B} = P_{n-1}^{DG}$, where P_n^C is the order n continuous Galerkin space, and P_{n-1}^{DG} is the order $n-1$ discontinuous Galerkin space. This yields the $Q_r^- \Lambda^k$ family from finite element exterior calculus. An analysis of the dispersion properties for the $P_1^C - P_0^{DG}$ and $P_2^C - P_1^{DG}$ pairs (in both 1D and 2D) can be found in [17, 18], including the use of partial lumping of the velocity mass matrix to remove spectral gaps present at $n \geq 2$. The 1D analysis for arbitrary n is presented below, including the use of both exact and inexact quadrature, along with a more general mass lumping procedure that attempts to remove spectral gaps for $n = 2$ and $n = 3$.

The degrees of freedom and basis functions associated with the $P_n^C - P_{n-1}^{DG}$ pair for $n = 2$ on a periodic mesh with $N = 4$ and $h = 1$ are shown in

Figures 1 and 2. For the $P_n^C - P_{n-1}^{DG}$ pair $l = n$, and unless otherwise noted, a uniform distribution of nodal points within an element is used. In practice, at higher orders, Gauss-Lobatto-Legendre (GLL) points are preferable due to the superior conditioning of the resulting matrices. However, when using exact quadrature the dispersion relationship is invariant to the choice of basis.

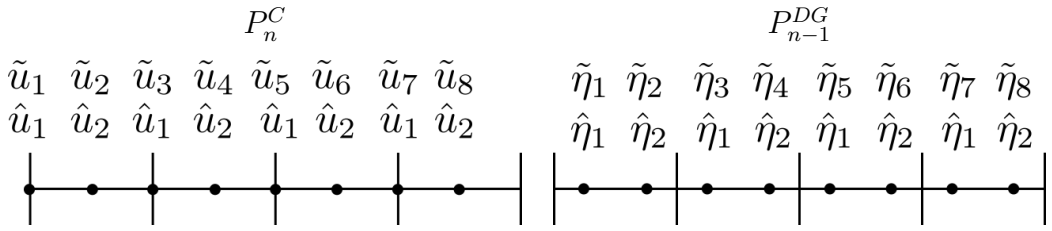

Figure 1: The degrees of freedom for the P_n^C pair (left) and P_{n-1}^{DG} pair (right) on a periodic mesh with $N = 4$, $h = 1$ and $n = 2$.

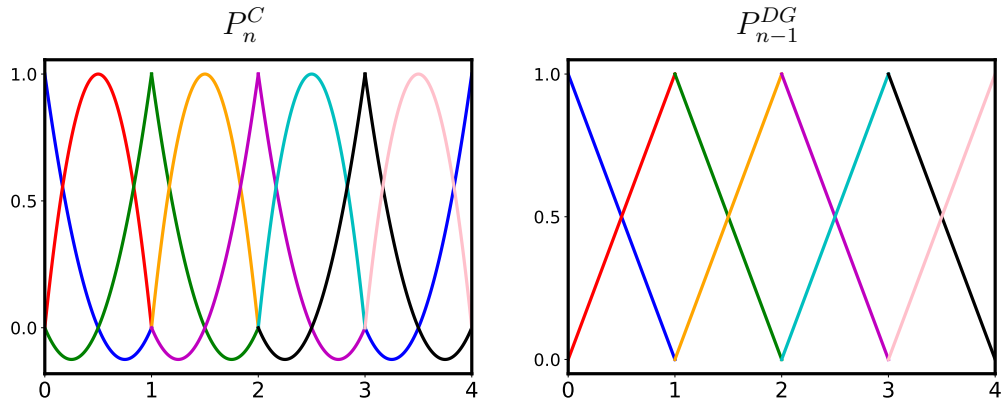

Figure 2: Basis functions for the P_n^C pair (left) and P_{n-1}^{DG} pair (right) on a periodic mesh with $n = 2$, $N = 4$ and $h = 1$.

4.1. Example for $n = 2$

As an example of the procedure described in Section 3, the construction of the matrix $\widehat{\mathbf{M}}_{\mathbf{u}}$ for $n = 2$ and the branch selection process are both illustrated. Start with a periodic grid consisting of N uniform elements of width h. The

$2N$ degrees of freedom for the H^1 space are located at

$$x_i = \left[0, \frac{h}{2}, \frac{2h}{2}, \ldots, \frac{(2N-1)h}{2}\right], \tag{51}$$

and

$$\tilde{\mathbf{u}} = [\tilde{u}_1, \ldots, \tilde{u}_{2N}] = [\hat{u}_1, \hat{u}_2, \hat{u}_1, \hat{u}_2, \ldots, \hat{u}_1, \hat{u}_2]\,. \tag{52}$$

Due to the translational invariance, the fundamental degrees of freedom are

$$\hat{\mathbf{u}} = [\hat{u}_1, \hat{u}_2]\,. \tag{53}$$

Considering a single element (for example, the one at $\frac{N}{2}$) the relevant rows of $\mathbf{M}_u$ are:

$$\mathbf{M_u} = \begin{bmatrix} -\frac{h}{30} & \frac{h}{15} & \frac{4h}{15} & \frac{h}{15} & -\frac{h}{30} \\ 0 & 0 & \frac{h}{15} & \frac{8h}{15} & \frac{h}{15} \end{bmatrix}. \tag{54}$$

By using (51), $\widetilde{\mathbf{M}}_\mathbf{u}$ is computed employing $(\widetilde{\mathbf{M}}_\mathbf{u})_{mn} = (\mathbf{M_u})_{mn} e^{ik(x_m - x_n)}$, and the relevant rows are:

$$\widetilde{\mathbf{M}}_\mathbf{u} = \begin{bmatrix} -\frac{h}{30}e^{ikh} & \frac{h}{15}e^{i\frac{kh}{2}} & \frac{4h}{15} & \frac{h}{15}e^{-i\frac{kh}{2}} & -\frac{h}{30}e^{-ikh} \\ 0 & 0 & \frac{h}{15}e^{i\frac{kh}{2}} & \frac{8h}{15} & \frac{h}{15}e^{-i\frac{kh}{2}} \end{bmatrix}. \tag{55}$$

Finally, $\widehat{\mathbf{M}}_\mathbf{u}$ is computed with $(\widehat{\mathbf{M}}_\mathbf{u})_{mn} = \sum_p (\widetilde{\mathbf{M}}_\mathbf{u})_{mp}$ and this leads to

$$\widehat{\mathbf{M}}_\mathbf{u} = \begin{bmatrix} \left(-\frac{h}{30}e^{2ikh} + \frac{4h}{15}e^{ikh} - \frac{h}{30}\right)e^{-ikh} & \left(\frac{h}{15}e^{\frac{3i}{2}kh} + \frac{h}{15}e^{\frac{ikh}{2}}\right)e^{-ikh} \\ \left(\frac{h}{15}e^{2ikh} + \frac{h}{15}e^{ikh}\right)e^{-\frac{3i}{2}kh} & \frac{8h}{15} \end{bmatrix}. \tag{56}$$

A similar construction holds for all the matrices needed to define the generalized eigenvalue problem (35) for ω^2.

Branch Selection. Once a set of n eigenvalues ω_n for a given kh have been obtained, the corresponding eigenvectors are constructed according to

$$u(x) = \sum_{j=1}^{Nl} \hat{u}_j e^{ikx_j} \psi_j(x), \tag{57}$$

where $\hat{u}_j$ is the elementwise eigenvector associated to the eigenvalue, and ψ_j is the basis function for the degree of freedom j. An example of this for $n = 2$ on a periodic mesh with $N = 12$ and $h = 1$ for $\widetilde{kh} = \frac{\pi}{4}$ is shown in Figure 3. It is clear in this case that the physical eigenvector is the low-frequency one. The same sort of analysis can be done for other values of n and $\widetilde{kh}$, and serves as a heuristic justification for the mode association procedure outlined in Section 3.4.

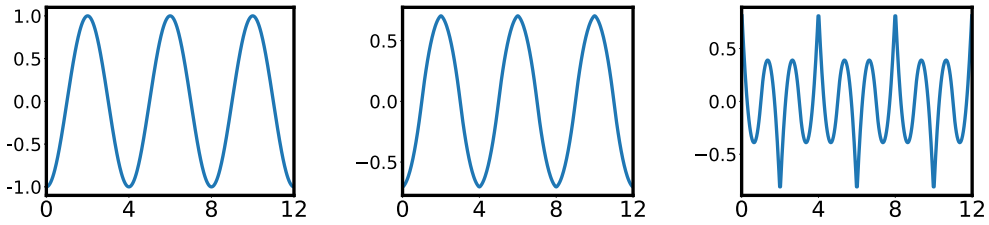

Figure 3: Eigenvectors for $n = 2$ and $k = \frac{\pi}{2h}$ (which corresponds to $\zeta = 4h = 8\tilde{h}$, $\tilde{\zeta} = 8$ and $\widetilde{kh} = \frac{\pi}{4}$) on a periodic mesh with 12 elements and $h = 1$. The left figure is the analytic eigenvector, the middle figure is the eigenvector for the low-frequency solution and the right figure is the eigenvector for the high-frequency solution. As expected, the eigenvector associated with the low-frequency solution corresponds to the analytic eigenvector, while the eigenvector associated with the high-frequency solution has high spatial wavenumber. This indicates that the low-frequency solution is the physical one, while the high-frequency solution is a mathematical artifact.

4.2. Exact Quadrature Asymptotics

On a uniform grid the highest degree polynomials that occur in the integrals are degree $2n$ (which occur only in $\mathbf{M_u}$), so by using a quadrature rule of order $2n$ or greater the integrals will be computed exactly. The canonical choice is Gauss-Lobatto quadrature (GL) using $n+1$ points, which is order $2n+1$. Exact quadrature yields the following results for the asymptotics

Theorem 1. *With exact quadrature, in the limit as mesh spacing $h \to 0$ we obtain the asymptotic results for $n \leq 15$:*

$$\omega \;=\; \omega_{AN} \pm \frac{1}{2^{2n+1}\displaystyle\prod_{j=1}^{n}(4j^2-1)} \frac{(-f^2 + g\,H\,k^2)}{\sqrt{f^2 + g\,H\,k^2}} (kh)^{2n} + O(h^{2n+2}).$$

It is conjectured that this result holds $\forall n \geq 1$.

These results were obtained using Maple and the analytic solution procedure described above, and are at the limits of the software. An attempt to go beyond $n = 15$ was made, but failed to complete within 1 week of computation time.

4.3. Inexact Quadrature Asymptotics

The use of exact quadrature leads to block diagonal $\mathbf{M_h}$ and $\mathbf{M_v}$, but not $\mathbf{M_u}$. Consequently, a global linear system has to be solved, even in the case of explicit time stepping schemes. The solution of such a linear system is often the dominant computational cost in a simulation. Therefore it is useful to consider choices of quadrature and basis that lead to diagonal mass matrices for all three variables, and the canonical choice is using $n+1$ GLL points for both basis and quadrature. When performed in the context of continuous Galerkin elements, such a procedure is known as the spectral element method [18]. Here we term it inexact quadrature. This choice of quadrature has an order of $2n-1$, and therefore on a uniform grid the matrix $\mathbf{M_u}$ will be underintegrated while the other matrices are not modified. However, $\mathbf{M_u}$ will now be diagonal, and no linear solver is required for explicit time stepping. Additionally, as shown in Appendix A, none of the desirable properties of the $P_n^C - P_{n-1}^{DG}$ pair will be affected by the use of inexact quadrature, even on a non-uniform grid with the nonlinear equations, provided it is used consistently for all integrals. The use of inexact quadrature yields the following results for the asymptotics

Theorem 2. *Using inexact quadrature, in the limit as mesh spacing $h \to 0$ we obtain the asymptotic results for $n \leq 6$*

$$\omega \;=\; \omega_{AN} \mp \frac{1}{2^{2n+1} n \prod_{j=1}^{n} (4j^2-1)} \frac{((2n+1)f^2 + g\,H\,k^2)}{\sqrt{f^2 + g\,H\,k^2}} (kh)^{2n} + O(h^{2n+2}).$$

It is conjectured that this result holds $\forall n \geq 1$.

The use of inexact quadrature has not affected the order of the dispersion relationship, only the sign and magnitude of the leading error term. These results were obtained using Maple and the analytic solution procedure described above. Unfortunately, to the authors knowledge, there do not exist closed form solutions for GLL quadrature points and weights for more than 7 points, and so it was not possible to obtain closed-form asympototic results past $n = 6$.

4.4. Dispersion Relationships and Group Velocities

Plots of $P_n(\sigma)$ versus $\widetilde{kh}$, the dispersion relationship $\sigma = \frac{\omega}{f}$ versus $\widetilde{kh}$ and the group velocity $d\sigma/d(\widetilde{kh})$ versus $\widetilde{kh}$ for the case of exact quadrature

with $\lambda/\tilde{h} = 2.0$ (a well-resolved Rossby radius) and $n = 1, \ldots, 6$, are found in Figure 4. By defining the Rossby radius via $\lambda/\tilde{h}$, the same effective resolution of the Rossby radius occurs for all n. When $n \geq 2$, spectral gaps are found at $\widetilde{kh} = \frac{j\pi}{n}$ $(j = 1, \ldots, n-1)$, where the dispersion relationship is double-valued and the group velocity goes to zero. For increasing values of n, the number of gaps increases and the size of final gap (the jump in frequency at the discontinuity) gets larger. This suggests that numerical dissipation will be required to control the spurious behavior in the short wavelength part of the spectrum. Indeed, it can be undesirable to introduce dissipation to control numerics rather than on physical grounds. However, in the long wavelength part of the spectrum the gaps decrease in size as n increases, and at the sampling frequency used ($N = \frac{6000}{n}$), they are not detectable. For a domain of width $L = 6371$km (corresponding to the equatorial radius of the Earth), this gives $\tilde{h} = 1.062$km. Therefore, it is not clear that these gaps will play a significant role in realistic simulations. This issue is explored in more detail in Section 6. On top of the gaps, there is also significant overestimation of the group velocity in the short wavelength part of the spectrum as n increases. Additionally, the maximum frequency increases with n, which will lead to more stringent CFL limits for explicit time stepping schemes. This seems to be a likely cause for the observation that maximum allowable time step decreases as a function of n for a fixed number of degrees of freedom.

Similar plots for the case of $\lambda/\tilde{h} = 0.1$ (a poorly resolved Rossby radius) are shown in Figure 5. The features are mostly the same as in the well-resolved case, with the exception that the last branch (the short wavelength part of the spectrum) now significantly underestimates the frequency and has a group velocity with the wrong sign. Using inexact quadrature, as shown in Figure 6 and Figure 7, does not materially alter these conclusions. The frequency is underestimated relative to the exact quadrature case, but spectral gaps are still present and the maximum frequency still increases with n (although less than in the exact quadrature case).

4.5. Lumping

As clearly demonstrated in the previous subsections, the $P_n^C - P_{n-1}^{DG}$ pair suffers from the presence of spectral gaps when $n \geq 2$. As shown in [17, 18] for $n = 2$, one way to eliminate these gaps is by partially lumping the $\mathbf{M_u}$ matrix. In Appendix A we show that such a lumping does not affect the desirable properties of the scheme, including energy conservation. Here we attempt to extend the lumping approach to the case $n = 3$.

4.5.1. The case $n = 2$

We start with the case $n = 2$. The element-wise $\mathbf{ML_u}$ matrix is lumped by introducing an auxiliary matrix

$$\mathbf{ML_a} = \begin{pmatrix} \alpha & \beta & -\alpha \\ \beta & \gamma & \beta \\ -\alpha & \beta & \alpha \end{pmatrix}, \tag{58}$$

designed to preserve the symmetry of $\mathbf{ML_u}$ and letting $\mathbf{ML_{u,new}} = \mathbf{ML_{u,old}} + \mathbf{ML_a}$, where $\mathbf{ML_{u,old}}$ is the element-wise mass matrix without lumping. Introducing the requirement that each row in $\mathbf{ML_a}$ sums to 0, to preserve positive-definiteness, immediately yields $\beta = \gamma = 0$. The spectral gap is eliminated by enforcing that the two branches of the dispersion relationship are equal at $\widetilde{kh} = \frac{\pi}{2}$. Computing the analytic solutions with $\mathbf{ML_{u,new}}$ and enforcing equality of the branches yields finally $\alpha = 1/30$. As might be expected, this reproduces the symmetric lumping from [17, 18]. Now applying the same procedure as before to compute asymptotics, we obtain

$$\omega = \omega_{AN} \mp \frac{\alpha}{4}\sqrt{f^2 + g\,H\,k^2}(kh)^2 + O(h^4).$$

Although the lumping has eliminated the spectral gap, the order of the dispersion relationship has dropped from 4th order to 2nd order.

4.5.2. The case $n = 3$

A similar approach is applied for $n = 3$, and we now consider the auxiliary matrix

$$\mathbf{ML_a} = \begin{pmatrix} \alpha & \beta & -\beta & -\alpha \\ \beta & \gamma & -\gamma & -\beta \\ -\beta & -\gamma & \gamma & \beta \\ -\alpha & -\beta & \beta & \alpha \end{pmatrix}, \tag{59}$$

which again preserves the symmetric positive-definiteness of $\mathbf{M_u}$. Now there are two spectral gaps that must be eliminated: one at $\widetilde{kh} = \frac{\pi}{3}$ involving the two lowest frequency branches, and one at $\widetilde{kh} = \frac{2\pi}{3}$ involving the two highest frequency branches. Enforcing equality of the branches yields

$$\gamma = \frac{729}{40}\left(\mu - \frac{1}{140}\right), \tag{60}$$

and additional equations (not shown) for α and β in terms of γ and μ. Such a lumping is not useful because the lumping parameters depend on

μ and realistic models do not have a single Rossby radius of deformation. Interestingly, this dependence on μ goes away when considering the gravity wave limit $f = 0$, highlighting the importance of using the full equation set. The inability to eliminate the spectral gaps for $n \geq 3$, and the significant loss in convergence for $n = 2$ when the gaps are eliminated motivates the use of an alternative choice of spaces for $\mathbb{A}$ and $\mathbb{B}$. This is described in the next section.

5. Mimetic Galerkin differences: the $GD_n - DGD_{n-1}$ pair

A set of compatible spaces that does not suffer from spectral gaps is given by $\mathbb{A} = GD_n$ and $\mathbb{B} = DGD_{n-1}$. The space GD_n is the space of order n Galerkin differences [22]. From this H_1 space, which is a partition of unity, a corresponding L_2 space can be constructed following the approach proposed in [3]. The basis for the L_2 space, called DGD_{n-1}, is defined in terms of the basis for the H^1 space. Historically, the $GD_3 - DGD_2$ pair was proposed independently by Dubos and Kritsikis ([23]) using a different approach based on finite volume fluxes and reconstructions prior to the development of Galerkin differences and the more general arbitrary order $GD_n - DGD_{n-1}$ pair. The $GD_n - DGD_{n-1}$ pair can be constructed for arbitrary odd order, $n = 1, 3, 5, \ldots$, and for $n = 1$ it is identical to the $P_1^C - P_0^{DG}$ pair. The major difference between the $GD_n - DGD_{n-1}$ and $P_n^C - P_{n-1}^{DG}$ pairs, is that higher order for the $GD_n - DGD_{n-1}$ pair is obtained by increasing the support of the basis functions beyond just neighboring elements, rather than adding additional, purely local degrees of freedom inside of an element as for the $P_n^C - P_{n-1}^{DG}$ pair. Therefore, the $GD_n - DGD_{n-1}$ pair has only one degree of freedom per geometric entity (independent of n), which in 1D are simply vertices and elements. The degrees of freedom associated with the $GD_n - DGD_{n-1}$ pair are illustrated in Figure 8. For the $GD_n - DGD_{n-1}$ pair $l = 1$, independent of order, and therefore $\tilde{h} = h$. An example of the basis functions for $N = 4$ and $n = 3$ are shown in Figure 9. In contrast to the $P_n^C - P_{n-1}^{DG}$ pair, each degree of freedom has now the same basis function at any order n.

5.1. Example for $n = 3$

As an example of the procedure described in Section 3, the construction of the $\widehat{\mathbf{M}}_{\mathbf{u}}$ matrix for $n = 3$ is illustrated. Consider a periodic grid with N

uniformly spaced elements of width h. The N degrees of freedom for $\tilde{\mathbf{u}}$ are located at nodes

$$x_j = (j-1)h, \quad j = 1, 2, 3, \cdots, N, \tag{61}$$

and are denoted by

$$\tilde{\mathbf{u}} = [\tilde{u}_1, \ldots, \tilde{u}_N] = [\hat{u}_1, \hat{u}_1, \ldots, \hat{u}_1]\,. \tag{62}$$

Using translational invariance, the fundamental degrees of freedom are

$$\hat{\mathbf{u}} = [\hat{u}_1]\,. \tag{63}$$

Considering now a single element (for example, the $\frac{N}{2}$ element) the relevant row of $\mathbf{M_u}$ is:

$$\mathbf{M_u} = \tfrac{h}{35}\left[\tfrac{31}{432} \quad -\tfrac{3}{2} \quad \tfrac{257}{48} \quad \tfrac{733}{27} \quad \tfrac{257}{48} \quad -\tfrac{3}{2} \quad \tfrac{31}{432}\right]. \tag{64}$$

By using (61), $\widetilde{\mathbf{M}}_\mathbf{u}$ is constructed employing $(\widetilde{\mathbf{M}}_\mathbf{u})_{mn} = (\mathbf{M_u})_{mn} e^{ik(x_m - x_n)}$, and the relevant row is:

$$\widetilde{\mathbf{M}}_\mathbf{u} = \tfrac{h}{35}\left[\tfrac{31}{432}e^{3ikh} \quad -\tfrac{3}{2}e^{2ikh} \quad \tfrac{257}{48}e^{ikh} \quad \tfrac{733}{27} \quad \tfrac{257}{48}e^{-ikh} \quad -\tfrac{3}{2}e^{-2ikh} \quad \tfrac{31}{432}e^{-3ikh}\right]. \tag{65}$$

Finally, $\widehat{\mathbf{M}}_\mathbf{u}$ is constructed using $(\widehat{\mathbf{M}}_\mathbf{u})_{mn} = \sum_p (\widetilde{\mathbf{M}}_\mathbf{u})_{mp}$, which yields

$$\widehat{\mathbf{M}}_\mathbf{u} = \left[\left(\tfrac{31h}{15120}e^{6ikh} - \tfrac{3h}{70}e^{5ikh} + \tfrac{257h}{1680}e^{4ikh} + \tfrac{733h}{945}e^{3ikh} + \tfrac{257h}{1680}e^{2ikh} - \tfrac{3h}{70}e^{ikh} + \tfrac{31h}{15120}\right)e^{-3ikh}\right]. \tag{66}$$

A similar construction holds for the other matrices needed to defined the generalized eigenvalue problem (35) for ω^2.

5.2. Exact Quadrature Asymptotics

As for the $P_n^C - P_{n-1}^{DG}$ pair, on a uniform grid the highest degree polynomials that occur in the integrals are of degree $2n$ (occurring only in $\mathbf{M_u}$). Therefore GL quadrature using $n+1$ points is exact. The use of exact quadrature yields the following results for the asymptotics:

Theorem 3. *Using exact quadrature, in the limit as mesh spacing $h \to 0$ we obtain the asymptotic results for $n \leq 9$:*

$$\omega = \omega_{AN} + \frac{\alpha_e}{\beta_e}\frac{(-f^2 + g\,H\,k^2)}{\sqrt{f^2 + g\,H\,k^2}}(kh)^{2n} + O(h^{2n+2}). \tag{67}$$

where α_e and β_e are given in Table 1. It is conjectured that (67) holds for $n \geq 10$.

Table 1: The coefficients α_e and β_e for the $GD_n - DGD_{n-1}$ pair using exact quadrature.

n	α_e	β_e
$n=1$	1	$2^3 \times 3$
$n=3$	17	$2^6 \times 3^3 \times 7$
$n=5$	13×317	$2^{10} \times 3^5 \times 5^2 \times 11$
$n=7$	47×318749	$2^{13} \times 3^6 \times 5^3 \times 7^2 \times 11 \times 13$
$n=9$	44734915633	$2^{18} \times 3^8 \times 5^3 \times 7^3 \times 13 \times 17 \times 19$

The result in Theorem 3 is obtained using Maple and the analytic solution procedure described above. An attempt to go beyond $n = 9$ was made, but failed to complete within 1 week of computation time as the limits of the software were reached.

5.3. Reduced Quadrature Asymptotics

The $GD_n - DGD_{n-1}$ pair has similar computational cost and identical matrix sparsity patterns to odd-order compatible isogeometric analysis (IGA, [24, 25]) using splines of maximal continuity. The pair shares with IGA an increased computational cost per degree of freedom relative to an equivalent order $P_n^C - P_{n-1}^{DG}$ pair. Therefore, it is useful to investigate methods of reducing the cost. A promising area of research for IGA is the use of reduced quadrature rules [26]. We follow a similar approach and investigate the use of 2-pt Gaussian quadrature independent of n. A reduced quadrature does not exist for $n = 1$ since 2-pt Gaussian quadrature is exact in this case. As proven in Appendix A, reduced quadrature does not effect any of the desirable properties of the scheme. It will also not change the matrix sparsity pattern, it merely reduces the cost to compute each matrix entry. Using reduced quadrature yields the following result:

Theorem 4. *Using reduced quadrature, in the limit as mesh spacing $h \to 0$ we obtain the asymptotic results for $n \leq 9$:*

$$\omega = \omega_{AN} + \frac{\alpha_r}{\beta_r} \frac{g\,H\,k^2}{\sqrt{f^2 + g\,H\,k^2}} (kh)^{n+1} + O(h^{n+3}), \tag{68}$$

where α_r and β_r are given in Table 2. It is conjectured that (68) holds for $n \geq 10$.

Table 2: The coefficients α_r and β_r for the mimetic element and reduced quadrature.

n	α_r	β_r
$n=3$	-1	$2^3 \times 3^3 \times 5$
$n=5$	-5	$2^5 \times 3^4 \times 7$
$n=7$	-17	$2^8 \times 3^3 \times 5 \times 7$
$n=9$	-133741	$2^{10} \times 3^9 \times 5 \times 7 \times 11$

As for (67) in Theorem 3 the computational time limits of the software were reached beyond $n = 9$. Significantly, the dispersion relationship has dropped from $2n$ order to $n + 1$ order.

5.4. Dispersion Relationships and Group Velocities

Plots of the dispersion relationship σ versus $\widetilde{kh}$ and the group velocity $d\sigma/d(\widetilde{kh})$ versus $\widetilde{kh}$ for the case of exact quadrature with $\lambda/\tilde{h} = 2.0$ (a well-resolved Rossby radius) and $n = 1, \ldots, 7$, are found in Figure 10. Unlike the $P_n^C - P_{n-1}^{DG}$ pair, there are no spectral gaps and the group velocity does not go to zero except at the end of the spectrum (which is the expected CD/Coriolis mode, [5]). Although there is a slight overestimation of the group velocity at high wavenumber, it is much less than for the $P_n^C - P_{n-1}^{DG}$ pair. Additionally, the maximal frequency does not increase with n. Therefore, the maximum allowable time step is expected to be insensitive to the choice of n for a fixed number of degrees of freedom. The same invariance of maximal time step to order n was found in [22] for the GD_n element applied to the wave equation. Similar plots for the case of $\lambda/\tilde{h} = 0.1$ (a poorly resolved Rossby radius) are shown in Figure 11. The features are mostly the same as the high-resolution case, with the exception that the short wavelength part of the spectrum now significantly underestimates the frequency and has a group velocity with the wrong sign. However, the long wavelength part of the spectrum is still well represented, and the poorly represented part of the spectrum gets smaller as n increases. Using reduced quadrature, as shown in Figure 12, does not materially alter these conclusions. The frequency is underestimated relative to the exact quadrature case, but there are still no spectral gaps, the group velocity stays close to the continuous one and the maximum frequency is still insensitive to the value of n. Comparing Figure 5 and Figure 11, we see

that for the case of $\lambda/\tilde{h} = 0.1$ the the long wavelength part of the dispersion relationship for the $P_n^C - P_{n-1}^{DG}$ pair is somewhat better than that for the $GD_n - DGD_{n-1}$ pair, especially in the region of the second to last branch. However, this is offset by the presence of a significant spectral gap in the short wavelength part of the spectrum. Therefore, it is not clear which of the two pairs will perform better. In fact, it seems likely that both pairs will perform poorly, just as the C-grid finite difference model does when λ is poorly resolved.

6. Simulations

To highlight the issues caused by the presence of spectral gaps, two different simulations were performed. Before proceeding to these tests, a slight digression into effective resolution is proposed.

6.1. Effective Resolution

A plot of the fractional error $\frac{\omega_N - \omega_{AN}}{\omega_{AN}}$ in the dispersion relationship versus the non-dimensional wavelength $\tilde{\zeta}$, where ω_N is the numerical frequency and ω_{AN} is the continuous frequency, is found in Figure 13. The calculation is performed for $\lambda/\tilde{h} = 2$ using exact quadrature with $n = 1, \ldots, 6$, in the case of the $P_n^C - P_{n-1}^{DG}$ pair and $n = 1, 3, 5, 7$, in the case of the $GD_n - DGD_{n-1}$ pair. The results are similar for other choices of quadrature and $\lambda/\tilde{h}$ (not shown). In particular, reduced quadrature for the $GD_n - DGD_{n-1}$ pair does not have a deleterious effect. In fact, it actually leads to a slight increase in effective resolution (not shown). From such an observation, an effective resolution can be defined by selecting an acceptable error level (here 0.01 or 1% is used) and determining the smallest non-dimensional wavelength with errors at or below this level. The largest gains in effective resolution come from increasing the order from $n = 1$ to $n = 3$, resulting in an effective resolution of about $4\tilde{h}$ for both the $GD_n - DGD_{n-1}$ and $P_n^C - P_{n-1}^{DG}$ pairs, with marginal gains when $n > 3$. These results can be compared to those in [27, 28, 29] who found similar conclusions for a variety of finite-difference, finite-element and finite-volume schemes.

6.2. Test Cases

Two distinct test cases are used to evaluate the ability of the schemes in representing the numerical solution without the presence of the spectral gaps. Unless otherwise noted, the domain Ω is $[0, L]$ with $N = 120/l$ elements and

and $h = L/N$, which ensures 120 degrees of freedom independent of l. The parameters are: $L = 2$ m, $\eta_0 = 1$ m, $\alpha = 1/60$, $x_0 = L/2$, $g = 1$ ms^{-1}, $H = 1$ m, $f = 25$ s^{-1}, which gives $\lambda = 1/25$ m. The unusual values of the parameters were chosen to yield $\lambda/\tilde{h} = 2.4$, which is a well-resolved Rossby radius, and to correspond to the tests run in [11, 17]. Implicit midpoint time stepping was used with $\Delta t = 1/120$ s, and 100 steps were employed.

6.2.1. The unsupported test case

In the first test, a Gaussian surface-elevation field is used with

$$\eta = \eta_0 \exp\left[-\frac{(x - x_0)^2}{\alpha^2}\right], \tag{69}$$

along with $u = v = 0$. This test was run using the $P_n^C - P_{n-1}^{DG}$ pair with both exact and inexact quadrature for $n = 1, \ldots, 6$, and using the $GD_n - DGD_{n-1}$ pair with both exact and reduced quadrature for $n = 3, 5, 7$. In addition, a pair of high-resolution runs with exact quadrature for 1440 degrees of freedom and $\Delta t = 1/1440$ s (with 1200 time steps) were performed to generate reference solutions. For the high-resolution runs, $n = 6$ and $N = 240$ were used for the $P_n^C - P_{n-1}^{DG}$ pair, while $n = 9$ and $N = 1440$ was used for the $GD_n - DGD_{n-1}$ pair.

Plots of the results for the $P_n^C - P_{n-1}^{DG}$ pair with exact quadrature can be found in Figure 14, and for the $GD_n - DGD_{n-1}$ pair with exact quadrature in Figure 15; the high-resolution reference solutions are found in Figure 16. For $n = 1$ the two elements are the same, and they both get the main features of the solution correct: a pair of propagating inertia-gravity wave packets and the steady geostrophic mode in the center. For $n > 1$ and the $P_n^C - P_{n-1}^{DG}$ pair, the main features are still present but there is now significant noise, which gradually improves for higher n. Unlike the results from [19], the noise does not appear to be located around element boundaries, except perhaps for $n = 5$ and $n = 6$. In contrast, the $GD_n - DGD_{n-1}$ pair does not exhibit this noise at any n. A possible explanation for the behavior of the $P_n^C - P_{n-1}^{DG}$ pair is as follows. The initial condition is projecting onto a range of wavenumbers, including those of the spectral gaps. However, the majority of the energy is concentrated in the lower frequencies. As n increases, the size of the gaps in the low frequency modes decreases, and therefore less energy is present at the problematic modes and the noise decreases. The use of inexact quadrature for the $P_n^C - P_{n-1}^{DG}$ pair or reduced quadrature for the $GD_n - DGD_{n-1}$ pair did not appreciably change the results (not shown). The difference between

the behavior of the two elements motivates the second simulation, designed explicitly to have significant energy at the spectral gaps.

6.2.2. The supported test case

In the second test case the Gaussian height field is modulated by a sinusoidal function designed to create a flow field with significant energy at the non-dimensional wavenumber of the spectral gaps, and is given by

$$\eta = \eta_0 \exp\left[-\frac{(x-x_0)^2}{a^2}\right] \cos(k(x-x_0)) \tag{70}$$

where $k = j\pi/(n\tilde{h})$ is the gap frequency for $j = 1, \ldots, n-1$. The test was run for $n = 5$ for both the $P_n^C - P_{n-1}^{DG}$ and $GD_n - DGD_{n-1}$ pairs with exact and inexact/reduced quadrature and $j = 1, 2, 3, 4$.

As shown in Figure 18, the $GD_5 - DGD_4$ pair worked well for all j and did not produce any noise in the simulation. In contrast, for the $P_5^C - P_4^{DG}$ pair shown in Figure 17, there is noise for all j. Unlike the results in [19], we identified problems even at the first gap ($j = 1$). This noise appears to be localized to element boundaries, as indicated by the vertical line structures. As j increased the simulation became increasingly inaccurate and noise starts to dominate the signal. Therefore, despite the size of the gaps in the low-frequency part of the spectrum decreasing as n increases, it is still possible to generate issues when wave packets have significant energy at the gap frequency. This might produce issues when nonlinear terms or physics parameterizations are added, since they are likely to generate forcing near the grid scale. As in the previous test, the use of inexact quadrature for the $P_n^C - P_{n-1}^{DG}$ pair or reduced quadrature for the $GD_n - DGD_{n-1}$ pair did not appreciably change the results (not shown).

7. Concluding remarks

The $P_n^C - P_{n-1}^{DG}$ pair has many desirable properties for geophysical fluid flow simulation, such as energy conservation, steady geostrophic modes, absence of any spurious stationary modes (such as pressure modes or inertial modes, with the exception of the expected CD/Coriolis mode) and no spurious branches of the dispersion relationship. Unfortunately, as shown in this work, it suffers from spectral gaps, poor group velocity at high wavenumbers and increasing maximal frequency with n, all of which grow worse with

increasing n. The presence of spectral gaps leads to anomalous wave dispersion for wave packets with significant energy close to the frequency of the gaps, despite the size of all but the last gap decreasing as n increases, and the increasing maximal frequency is expected to cause a decrease in maximal time step for a fixed number of degrees of freedom. These problems could be ameliorated through the addition of numerical dissipation at the relevant scales, but this is undesirable from physical grounds. Inexact quadrature can be used to reduce the computational cost by diagonalizing the mass matrix, but does not change the salient features and behavior. It is also not clear how to extend inexact quadrature to multiple dimensions, since there is no longer a choice of quadrature that diagonalizes the velocity mass matrix. In addition, attempts to remove the spectral gaps via mass lumping for $n = 3$ failed, and for $n = 2$ lead to a significant loss of convergence in the dispersion relationship. It is not clear if other approaches to lumping might be viable, although the proposed lumping for $n = 3$ is quite general. In any case it seems extremely likely that any lumping that removes the spectral gaps will lead to decreased convergence in the dispersion relationship, based on the results obtained for $n = 2$.

However, the $P_n^C - P_{n-1}^{DG}$ pair is part of a general class of discretizations known as compatible Galerkin methods. Another member of this class, the $GD_n - DGD_{n-1}$ pair, has the same desirable properties but much better dispersion properties. In particular, for any n the $GD_n - DGD_{n-1}$ pair is free of spectral gaps and the maximal frequency does not increase with n. Unfortunately, it does have increased computational cost relative to the $P_n^C - P_{n-1}^{DG}$ pair. Reduced quadrature, in the form of 2-pt Gaussian quadrature independent of n, can significantly reduce this computational cost without changing any of the features of the dispersion relationship or other desirable properties. An exception is the loss in order of convergence from $2n$ to $n + 1$, which is still less than the loss seen in the mass lumped $P_2^C - P_1^{DG}$ pair. Plots of effective resolution indicate that $n = 3$ might be a sweet spot in terms of maximizing resolvable wavenumber and order of convergence while retaining acceptable computational cost, especially if using reduced quadrature. Therefore, it appears that the $GD_n - DGD_{n-1}$ pair represents a viable alternative to the $P_n^C - P_{n-1}^{DG}$ pair for the development of geophysical fluid flow models, and is worthy of future study. This work, along with an extension of this study to 2D inertia-gravity waves and with the incorporation of time discretization, is currently underway.

8. Acknowledgements

Christopher Eldred appreciates support from the French National Research Agency through contract ANR-14-CE23-0010 (HEAT). The authors appreciate the thorough and detailed reviews of Thomas Melvin and an anonymous reviewer, which have greatly improved the clarity and presentation of this work.

9. References

[1] C. Cotter, J. Shipton, Mixed finite elements for numerical weather prediction, Journal of Computational Physics 231 (21) (2012) 7076 – 7091. doi:https://doi.org/10.1016/j.jcp.2012.05.020.

[2] A. Natale, J. Shipton, C. J. Cotter, Compatible finite element spaces for geophysical fluid dynamics, Dynamics and Statistics of the Climate System 1 (1) (2016) dzw005. doi:10.1093/climsys/dzw005.

[3] R. Hiemstra, D. Toshniwal, R. Huijsmans, M. Gerritsma, High order geometric methods with exact conservation properties, Journal of Computational Physics 257 (Part B) (2014) 1444 – 1471, physics-compatible numerical methods. doi:https://doi.org/10.1016/j.jcp.2013.09.027.

[4] D. N. Arnold, R. S. Falk, R. Winther, Finite element exterior calculus, homological techniques, and applications, Acta Numerica 15 (2006) 1155. doi:10.1017/S0962492906210018.

[5] V. Rostand, D. Y. L. Roux, G. Carey, Kernel analysis of the discretized finite difference and finite element shallow-water models, SIAM Journal on Scientific Computing 31 (1) (2008) 531–556. doi:10.1137/070695198.

[6] V. Rostand, D. Y. Le Roux, Raviart-thomas and brezzi-douglas-marini finite-element approximations of the shallow-water equations, International Journal for Numerical Methods in Fluids 57 (8) (2008) 951–976. doi:10.1002/fld.1668.

[7] D. Y. L. Roux, Spurious inertial oscillations in shallow-water models, Journal of Computational Physics 231 (24) (2012) 7959 – 7987. doi:https://doi.org/10.1016/j.jcp.2012.04.052.

[8] A. Arakawa, V. R. Lamb, Computational design of the basic dynamical processes of the ucla general circulation model 17 (Supplement C) (1977) 173 – 265. doi:https://doi.org/10.1016/B978-0-12-460817-7.50009-4.

[9] D. Y. L. Roux, V. Rostand, B. Pouliot, Analysis of numerically induced oscillations in 2d finite-element shallow-water models part i: Inertia-gravity waves, SIAM Journal on Scientific Computing 29 (1) (2007) 331–360. doi:10.1137/060650106.

[10] D. Y. L. Roux, B. Pouliot, Analysis of numerically induced oscillations in two-dimensional finite-element shallow-water models part ii: Free planetary waves, SIAM Journal on Scientific Computing 30 (4) (2008) 1971–1991. doi:10.1137/070697872.

[11] T. Melvin, A. Staniforth, J. Thuburn, Dispersion analysis of the spectral element method, Quarterly Journal of the Royal Meteorological Society 138 (668) (2012) 1934–1947. doi:10.1002/qj.1906.

[12] M. Ainsworth, H. A. Wajid, Dispersive and dissipative behavior of the spectral element method, SIAM Journal on Numerical Analysis 47 (5) (2009) 3910–3937. doi:10.1137/080724976.

[13] M. Ainsworth, Dispersive behaviour of high order finite element schemes for the one-way wave equation, Journal of Computational Physics 259 (Supplement C) (2014) 1 – 10. doi:https://doi.org/10.1016/j.jcp.2013.11.003.

[14] J. Thuburn, C. J. Cotter, A primaldual mimetic finite element scheme for the rotating shallow water equations on polygonal spherical meshes, Journal of Computational Physics 290 (2015) 274 – 297. doi:https://doi.org/10.1016/j.jcp.2015.02.045.
URL http://www.sciencedirect.com/science/article/pii/S0021999115001151

[15] T. Melvin, J. Thuburn, Wave dispersion properties of compound finite elements, Journal of Computational Physics 338 (2017) 68 – 90. doi:https://doi.org/10.1016/j.jcp.2017.02.025.
URL http://www.sciencedirect.com/science/article/pii/S0021999117301183

[16] C. J. Cotter, D. A. Ham, C. C. Pain, A mixed discontinuous/continuous finite element pair for shallow-water

ocean modelling, Ocean Modelling 26 (1) (2009) 86 – 90. doi:https://doi.org/10.1016/j.ocemod.2008.09.002.

[17] A. Staniforth, T. Melvin, C. Cotter, Analysis of a mixed finite-element pair proposed for an atmospheric dynamical core, Quarterly Journal of the Royal Meteorological Society 139 (674) (2013) 1239–1254. doi:10.1002/qj.2028.

[18] T. Melvin, A. Staniforth, C. Cotter, A two-dimensional mixed finite-element pair on rectangles, Quarterly Journal of the Royal Meteorological Society 140 (680) (2014) 930–942. doi:10.1002/qj.2189.

[19] T. Melvin, Dispersion analysis of the pnpn1dg mixed finite element pair for atmospheric modelling, Journal of Computational Physics 355 (2018) 342 – 365. doi:https://doi.org/10.1016/j.jcp.2017.11.019.
URL http://www.sciencedirect.com/science/article/pii/S0021999117308574

[20] Waves in the ocean, Vol. 20 of Elsevier Oceanography Series, Elsevier, 1978. doi:https://doi.org/10.1016/S0422-9894(08)70810-9.

[21] P. D. Powell, Calculating Determinants of Block Matrices, arXiv:1112.4379.

[22] J. Banks, T. Hagstrom, On galerkin difference methods, Journal of Computational Physics 313 (Supplement C) (2016) 310 – 327. doi:https://doi.org/10.1016/j.jcp.2016.02.042.

[23] E. Kritsikis, T. Dubos, Higher-order finite elements for the shallow-water equations on the cubed sphere, in: PDEs on the Sphere Workshop, 2014.

[24] A. Buffa, J. Rivas, G. Sangalli, R. Vazquez, Isogeometric discrete differential forms in three dimensions, SIAM Journal on Numerical Analysis 49 (2) (2011) 818–844. doi:10.1137/100786708.

[25] A. Buffa, G. Sangalli, R. Vazquez, Isogeometric methods for computational electromagnetics: B-spline and t-spline discretizations, Journal of Computational Physics 257 (Part B) (2014) 1291 – 1320, physics-compatible numerical methods. doi:https://doi.org/10.1016/j.jcp.2013.08.015.

[26] A. Sarmiento, A. Cortes, D. Garcia, L. Dalcin, N. Collier, V. Calo, Petiga-mf: A multi-field high-performance toolbox for structure-preserving b-splines spaces, Journal of Computational Science 18 (Supplement C) (2017) 117 – 131. doi:https://doi.org/10.1016/j.jocs.2016.09.010.

[27] J. Kent, J. P. Whitehead, C. Jablonowski, R. B. Rood, Determining the effective resolution of advection schemes. part i: Dispersion analysis, Journal of Computational Physics 278 (Supplement C) (2014) 485 – 496. doi:https://doi.org/10.1016/j.jcp.2014.01.043.

[28] J. Kent, C. Jablonowski, J. P. Whitehead, R. B. Rood, Determining the effective resolution of advection schemes. part ii: Numerical testing, Journal of Computational Physics 278 (Supplement C) (2014) 497 – 508. doi:https://doi.org/10.1016/j.jcp.2014.08.045.

[29] P. A. Ullrich, Understanding the treatment of waves in atmospheric models. part 1: The shortest resolved waves of the 1d linearized shallow-water equations, Quarterly Journal of the Royal Meteorological Society 140 (682) (2014) 1426–1440. doi:10.1002/qj.2226.

Appendix A. Energy conservation for inexact/reduced quadrature and lumping

Start with the definitions for kinetic energy

$$KE = H\frac{\mathbf{u}^T\mathbf{M_u}\mathbf{u}}{2} + H\frac{\mathbf{v}^T\mathbf{M_v}\mathbf{v}}{2}, \tag{A.1}$$

and potential energy

$$PE = g\frac{\boldsymbol{\eta}^T\mathbf{M}_{\boldsymbol{\eta}}\boldsymbol{\eta}}{2}, \tag{A.2}$$

and the evolution equations

$$\mathbf{M_u}\frac{\partial \mathbf{u}}{\partial t} - f\mathbf{C_u}\mathbf{v} + g\mathbf{G}\boldsymbol{\eta} = 0, \tag{A.3}$$

$$\mathbf{M_v}\frac{\partial \mathbf{v}}{\partial t} + f\mathbf{C_v}\mathbf{u} = 0, \tag{A.4}$$

$$\mathbf{M}_{\boldsymbol{\eta}}\frac{\partial \boldsymbol{\eta}}{\partial t} + H\mathbf{D}\mathbf{u} = 0. \tag{A.5}$$

Multiply (A.3) by $H\mathbf{u}^T$, (A.4) by $H\mathbf{v}^T$ and (A.5) by $g\boldsymbol{\eta}^T$ to get

$$H\mathbf{u}^T\mathbf{M}_\mathbf{u}\frac{\partial \mathbf{u}}{\partial t} - fH\mathbf{u}^T\mathbf{C}_\mathbf{u}\mathbf{v} + gH\mathbf{u}^T\mathbf{G}\boldsymbol{\eta} = 0, \tag{A.6}$$

$$H\mathbf{v}^T\mathbf{M}_\mathbf{v}\frac{\partial \mathbf{v}}{\partial t} + fH\mathbf{v}^T\mathbf{C}_\mathbf{v}\mathbf{u} = 0, \tag{A.7}$$

$$g\boldsymbol{\eta}^T\mathbf{M}_{\boldsymbol{\eta}}\frac{\partial \boldsymbol{\eta}}{\partial t} + gH\boldsymbol{\eta}^T\mathbf{D}\mathbf{u} = 0. \tag{A.8}$$

Time differentiate (A.1) and (A.2) and combine with (A.6) - (A.8) to get the evolution of the total energy $TE = KE + PE$

$$\frac{\partial (KE + PE)}{\partial t} = -gH\left[\mathbf{u}^T\mathbf{G}\boldsymbol{\eta} + \boldsymbol{\eta}^T\mathbf{D}\mathbf{u}\right] - Hf\left[-\mathbf{u}^T\mathbf{C}_\mathbf{u}\mathbf{v} + \mathbf{v}^T\mathbf{C}_\mathbf{v}\mathbf{u}\right] = 0. \tag{A.9}$$

This derivation relies on the properties (17) - (20). Therefore, any inexact/reduced quadrature or mass lumping that preserves these properties will still conserve energy, albeit with a possibly modified definition of energy. This is the case for all of the inexact/reduced quadratures and mass lumpings considered in this paper. This also extends to the nonlinear equations in 2D, provided the same velocity mass matrix is used to diagnose the mass flux $\vec{\mathbf{F}} = \boldsymbol{\eta}\vec{\mathbf{v}}$, define the kinetic energy and compute $\partial\vec{\mathbf{v}}/\partial t$.

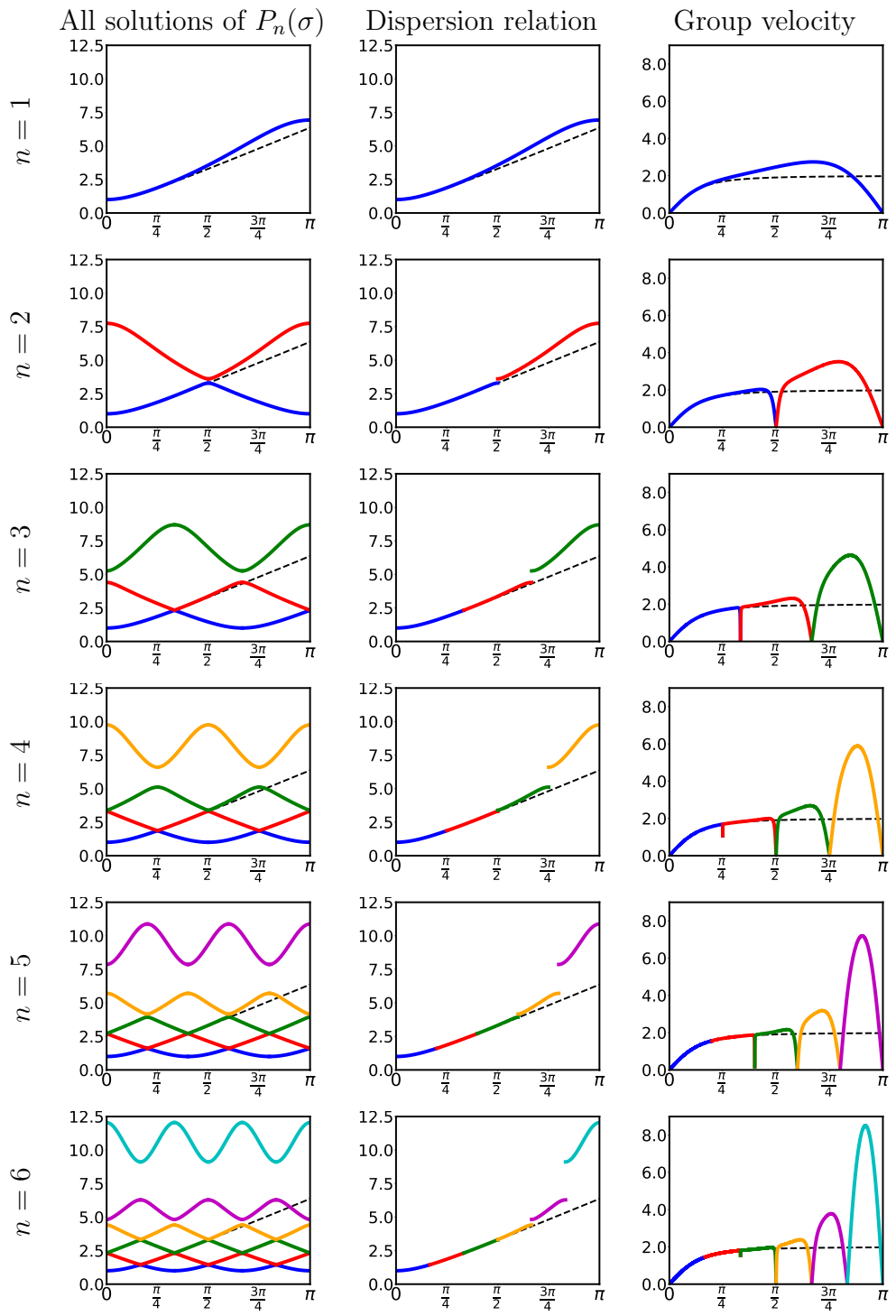

Figure 4: Dispersion relation and group velocity for the $P_n^C - P_{n-1}^{DG}$ pair using exact quadrature for $n = 1, \ldots, 6$, with $\lambda/\tilde{h} = 2$. The exact dispersion relation and group velocity are shown with dashed lines. The left column plots all solutions of the dispersion polynomial $P_n(\sigma)$, the middle column the physical branches, and the right column the group velocity $d\sigma/d(\widetilde{kh})$. Each solution and branch are given a distinct color. Spectral gaps are present for $n \geq 2$ and group velocity goes to zero at $\widetilde{kh} = (i\pi)/n$, where $i = 1, \ldots, n-1$. There is also significant overestimation of the group velocity for short wavelengths and maximal σ increases as n increases.

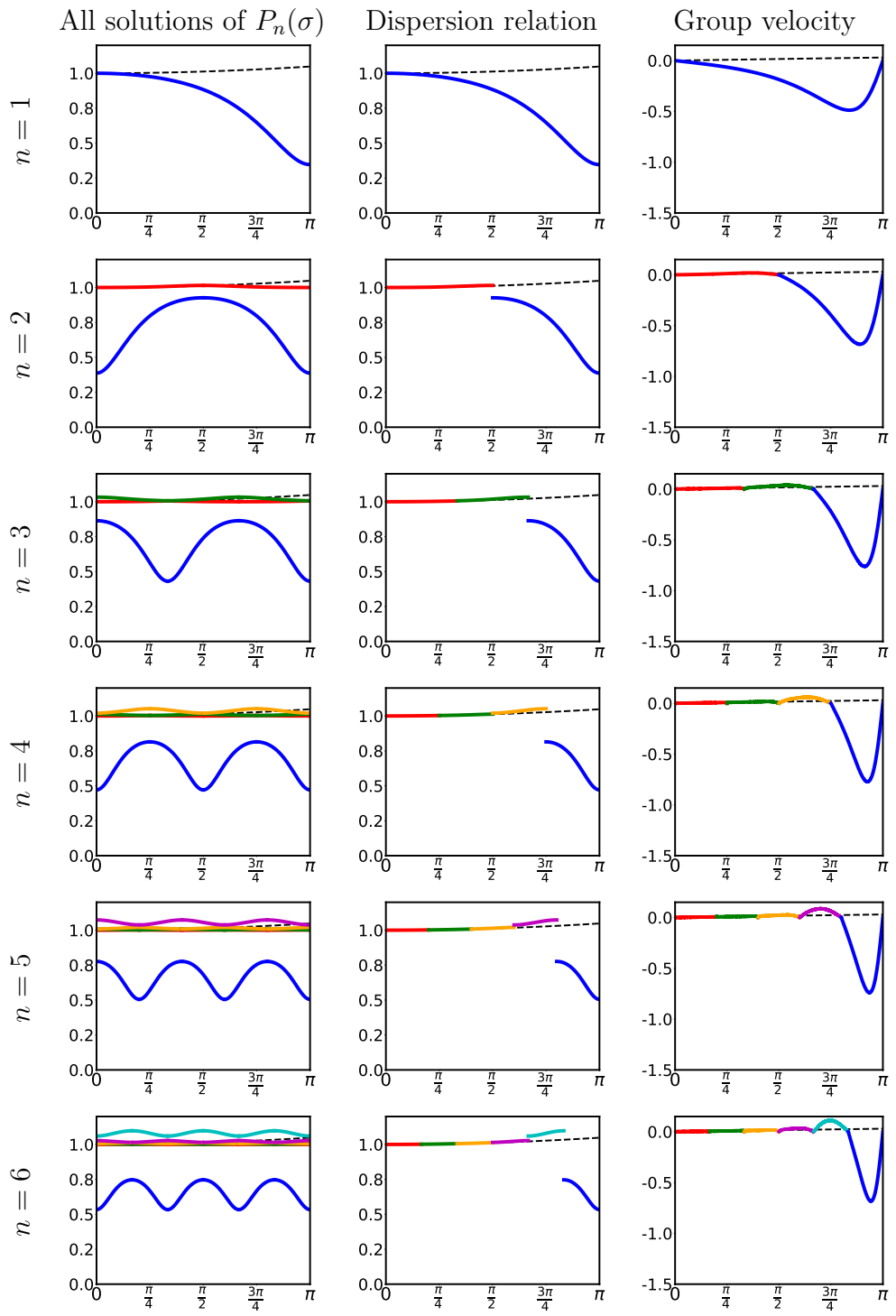

Figure 5: As for Figure 4 but with $\lambda/\tilde{h} = 0.1$. The main features are quite similar (spectral gaps, increasing maximal frequency with n), but the maximal frequency now occurs for the 2nd to last branch. The last branch significantly underestimates the frequency and has a group velocity with the wrong sign.

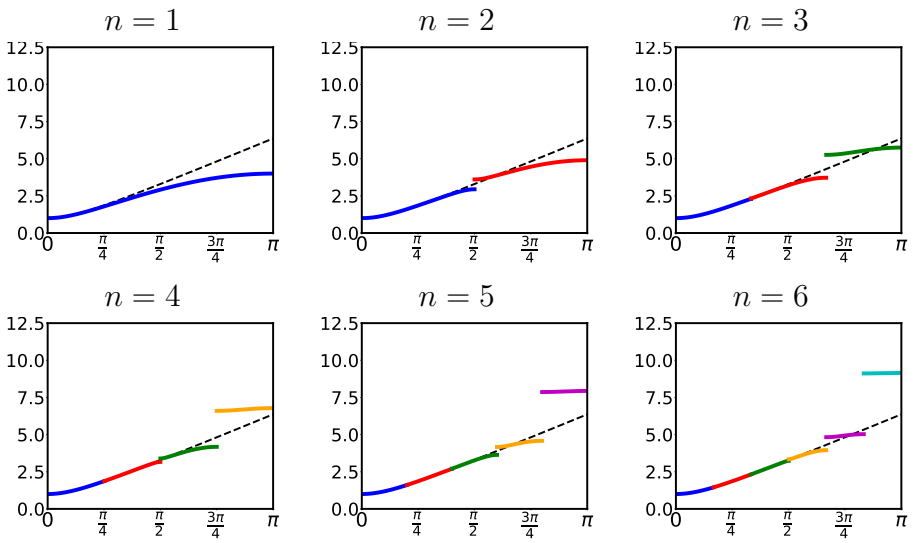

Figure 6: Dispersion relation for the $P_n^C - P_{n-1}^{DG}$ pair using inexact quadrature with $\lambda/\tilde{h} = 2.0$. The exact dispersion relation is shown with dashed lines. Although the frequency is now underestimated compared with exact quadrature, there are still spectral gaps and the maximum frequency still increases with n.

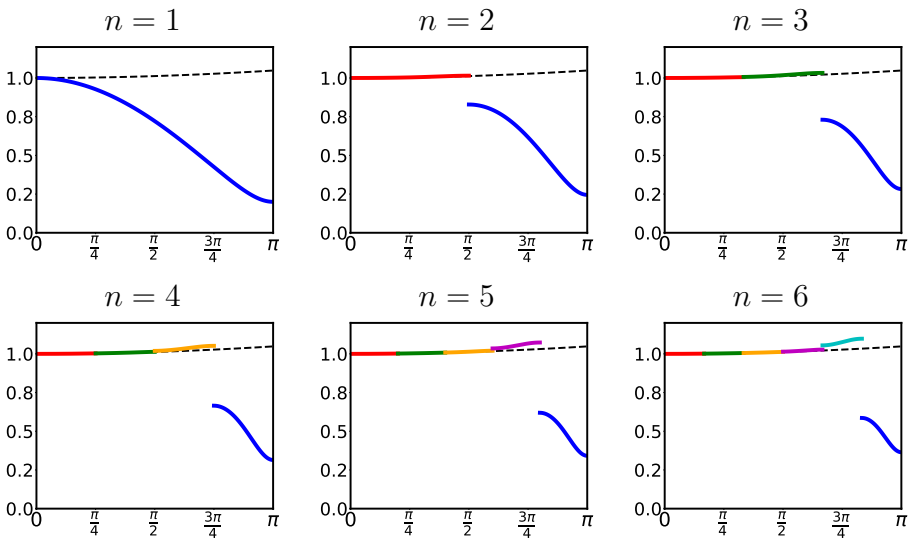

Figure 7: As for Figure 6 but with $\lambda/\tilde{h} = 0.1$. The key features have not changed from exact quadrature, the only major difference is that the frequency is underestimated relative to the case of exact quadrature. Spectral gaps, increasing maximum frequency and a significantly underestimated last branch with the wrong sign for group velocity are all still present.

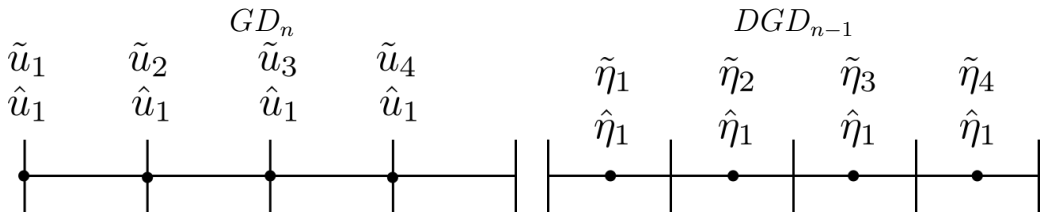

Figure 8: The degrees of freedom for the GD_n pair (left) and DGD_{n-1} pair (right) on a periodic mesh with $N = 4$ and $h = 1$.

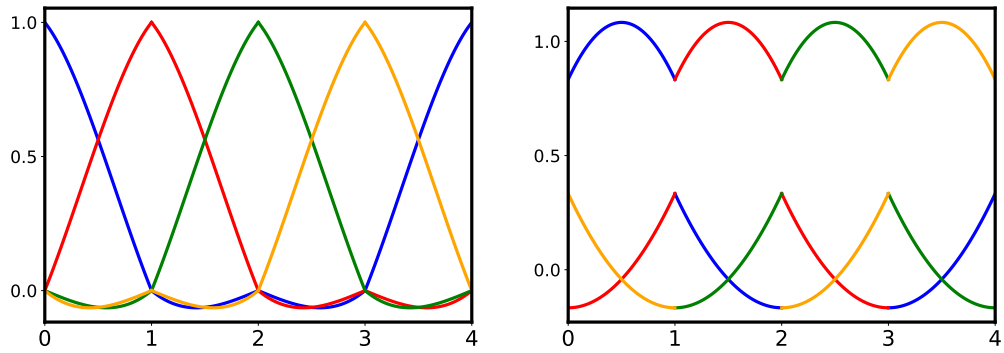

Figure 9: Basis functions for the GD_n pair (left) and DGD_{n-1} pair (right) on a periodic mesh with $N = 4$, $n = 3$ and $h = 1$.

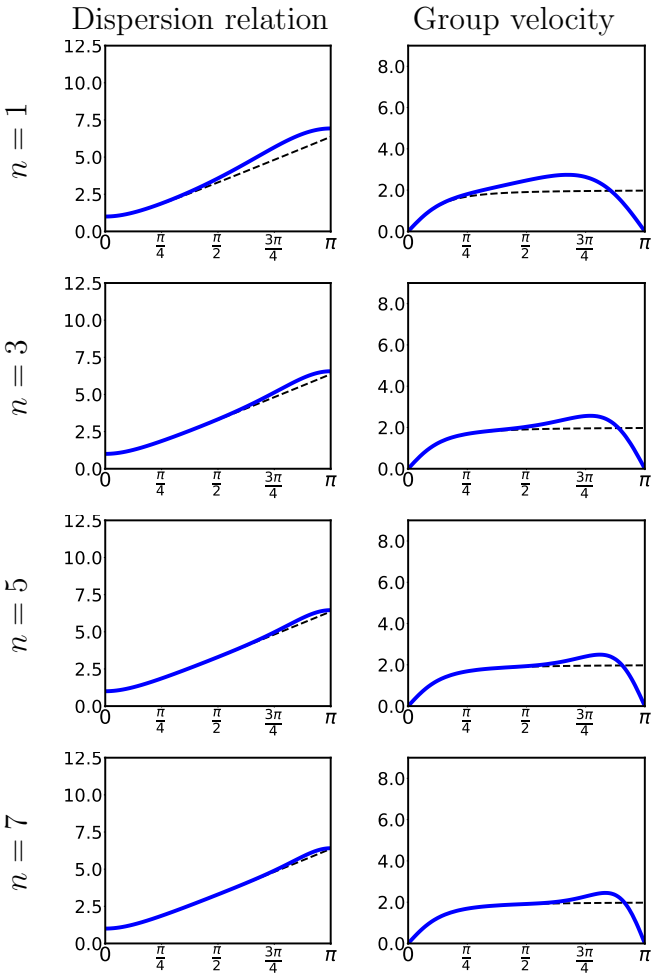

Figure 10: Dispersion relation and group velocity for the $GD_n - DGD_{n-1}$ pair at $n = 1, 3, 5, 7$ using exact quadrature with $\lambda/\tilde{h} = 2.0$. The exact dispersion relation and group velocity are shown with dashed lines. There are no spectral gaps, and there is a generally good approximation to both dispersion relation and group velocity with the exception of the CD/Coriolis mode at the end of the spectrum where group velocity goes to zero.

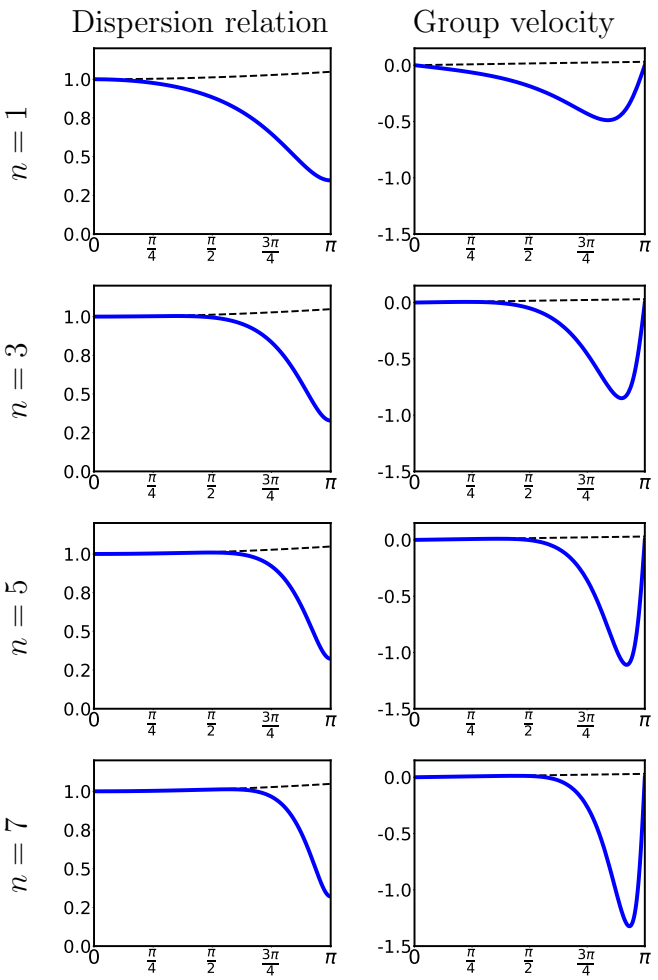

Figure 11: As for Figure 10 but with $\lambda/\tilde{h} = 0.1$. The exact dispersion relation and group velocity are shown with dashed lines. Just like the $\lambda/\tilde{h} = 2.0$ case, there are no spectral gaps but there is a CD/Coriolis mode. The dispersion relation and the group velocity are well-approximated in the long wavelength part of the spectrum, which gets bigger as n increases. However, as for the $P_n^C - P_{n-1}^{DG}$ pair, the high wavenumber part of the spectrum is still very poor.

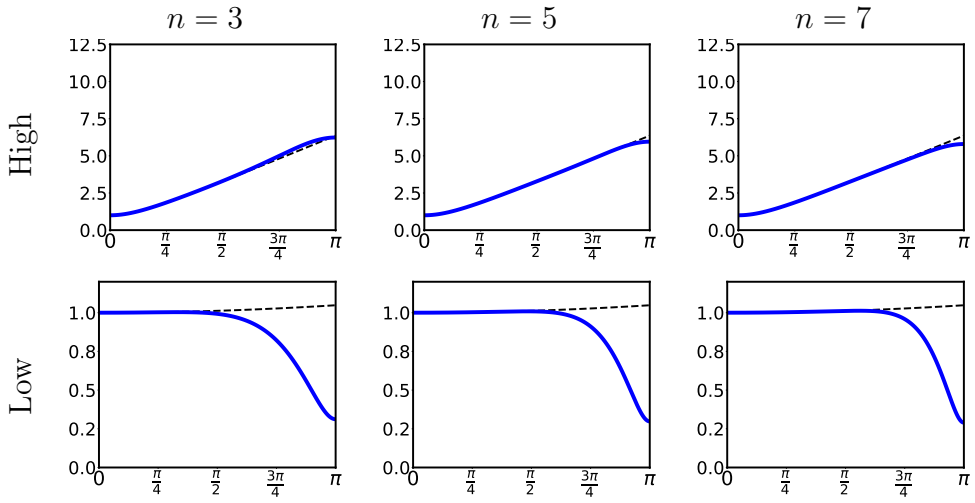

Figure 12: Dispersion relation for the $GD_n - DGD_{n-1}$ pair at $n = 3, 5, 7$ using reduced quadrature with $\lambda/\tilde{h} = 2.0$ (High) and $\lambda/\tilde{h} = 0.1$ (Low). The case $n = 1$ is not considered since 2-pt Gaussian quadrature is exact when $n = 1$. In general, the dispersion relation is underestimated in contrast to the exact case. However, it is still quite good for $\lambda/\tilde{h} = 2.0$, and there is still no increase in maximal frequency with increasing n or spectral gaps.

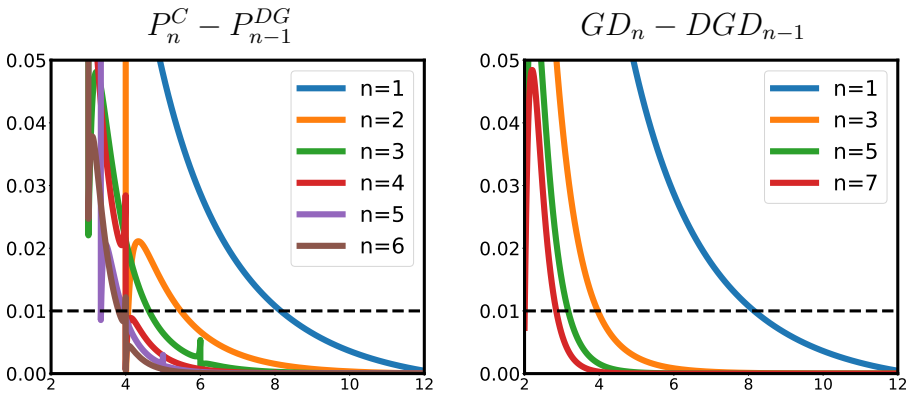

Figure 13: Fractional error $\frac{\omega_N - \omega_{AN}}{\omega_{AN}}$ for $\lambda/\tilde{h} = 2.0$ using exact quadrature with $n = 1, \ldots, 6$ for the $P_n^C - P_{n-1}^{DG}$ pair and $n = 1, 3, 5, 7$ for the $GD_n - DGD_{n-1}$ pair, as a function of $\tilde{\zeta}$. The horizontal dashed line indicates an error level of 0.01, which corresponds to a 1% error in the dispersion relationship. From this, an effective horizontal resolution can be defined (see text). The largest increases come from increasing order from 1 to 3, with diminishing returns past $n = 3$.

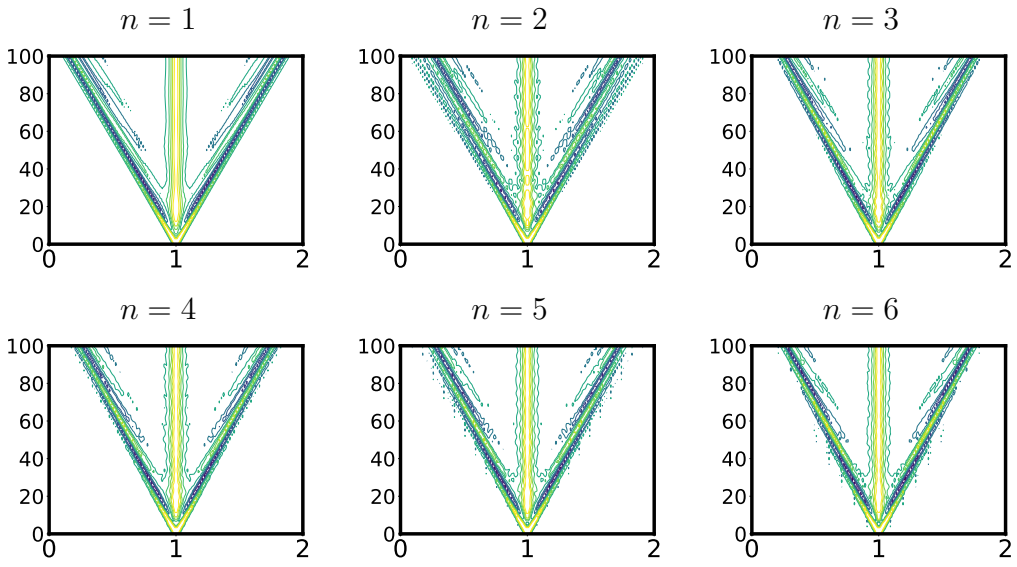

Figure 14: Hovmoller plots of η for the unsupported test case using the $P_n^C - P_{n-1}^{DG}$ pair with exact quadrature for $n = 1, \ldots, 6$. The x-axis is the spatial coordinate x, and the y-axis is time step. For $n > 1$, there is significant noise, which somewhat improves as n increases but does not go away. Unlike [19], the noise does not appear to be localized to element boundaries, except perhaps for $n = 5$ and $n = 6$. The contour interval is 0.05m, centered around 0.

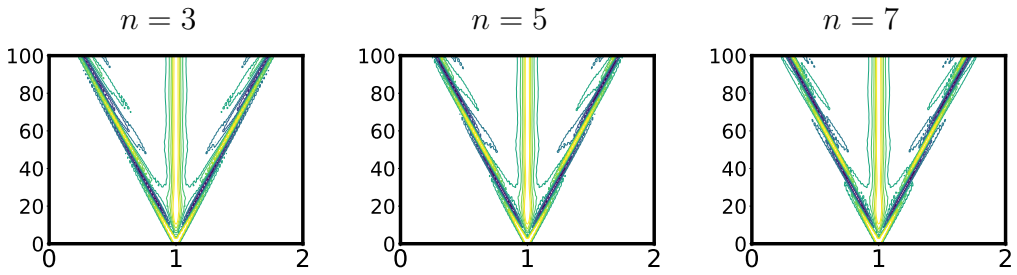

Figure 15: Hovmoller plots of η for the unsupported test case using the $GD_n - DGD_{n-1}$ pair with exact quadrature for $n = 3, 5, 7$. The x-axis is the spatial coordinate x, and the y-axis is time step. There is no noise for any n. The contour interval is 0.05m, centered around 0.

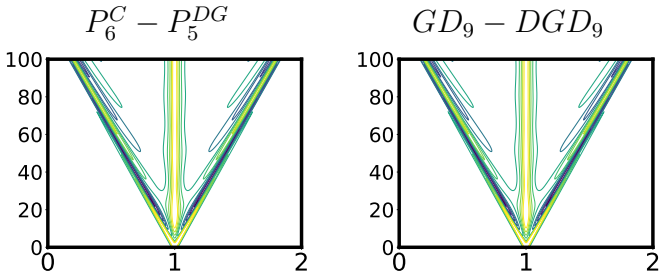

Figure 16: Hovmoller plots of η for the unsupported test case at high resolution using the $P_6^C - P_5^{DG}$ and $GD_9 - DGD_9$ pairs with exact quadrature. The x-axis is the spatial coordinate x, and the y-axis is time step. There is essentially no difference between the results, indicating that the majority of the energy is concentrated in the low-frequency part of the spectrum at this resolution. The contour interval is 0.05m, centered around 0.

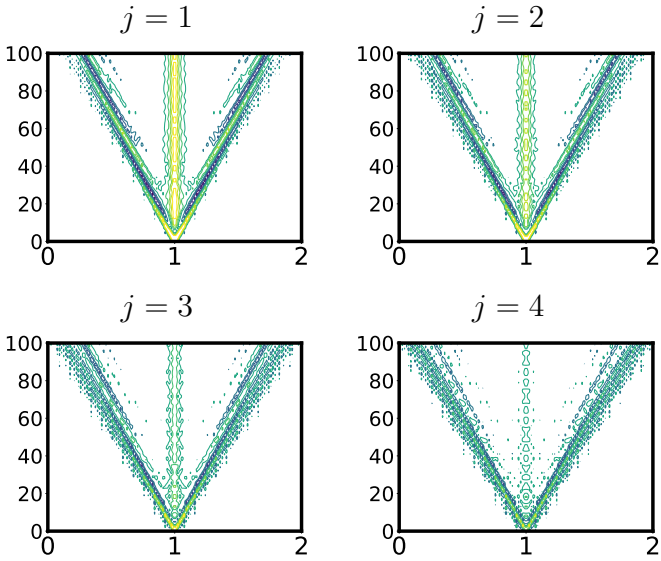

Figure 17: Hovmoller plot of η for the supported test case using the $P_5^C - P_4^{DG}$ pair with exact quadrature. The x-axis is the spatial coordinate x, and the y-axis is time step. There is noise for all values of j, with increasing noise as j increases. This indicates that the spectral gaps at higher j are easier to excite. Unlike the unsupported test case, the noise does appear to be localized to element boundaries, as indicated by the vertical line structures. The contour interval is 0.05m, centered around 0.

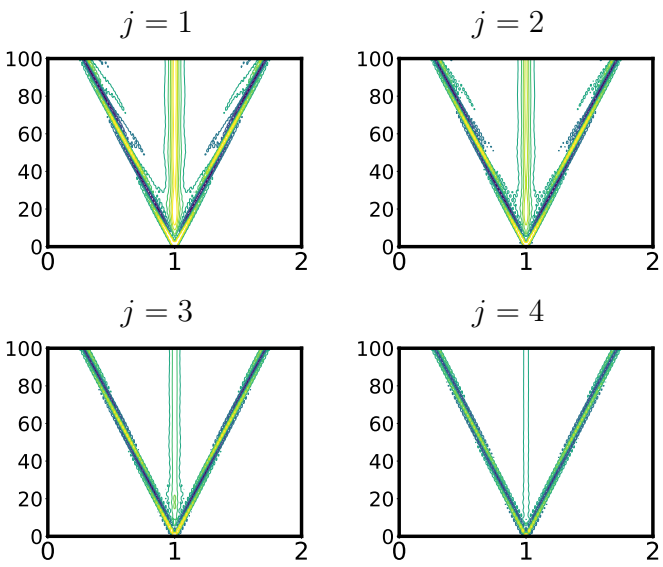

Figure 18: Hovmoller plot of η for the supported test case using the $GD_5 - DGD_4$ pair with exact quadrature. The x-axis is the spatial coordinate x, and the y-axis is time step. There is no noise at any j. The contour interval is 0.05m, centered around 0.